Target SQL Business Case

Context:

Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

Dataset: not provided for pubblic use

The data is available in 8 csv files:

1. customers.csv
2. sellers.csv
3. order_items.csv
4. geolocation.csv
5. payments.csv
6. reviews.csv
7. orders.csv
8. products.csv

The column description for these csv files is given below.

The **customers.csv** contain following features:

Features	Description
customer_id	ID of the consumer who made the purchase

customer_unique_id	Unique ID of the consumer
customer_zip_code_prefix	Zip Code of consumer's location
customer_city	Name of the City from where order is made
customer_state	State Code from where order is made (Eg. são paulo - SP)

The **sellers.csv** contains following features:

Features	**Description**
seller_id	Unique ID of the seller registered
seller_zip_code_prefix	Zip Code of the seller's location
seller_city	Name of the City of the seller
seller_state	State Code (Eg. são paulo - SP)

The **order_items.csv** contain following features:

Features	**Description**
order_id	A Unique ID of order made by the consumers
order_item_id	A Unique ID given to each item ordered in the order
product_id	A Unique ID given to each product available on the site
seller_id	Unique ID of the seller registered in Target
shipping_limit_date	The date before which the ordered product must be shipped
price	Actual price of the products ordered
freight_value	Price rate at which a product is delivered from one point to another

The **geolocations.csv** contain following features:

Features	**Description**
geolocation_zip_code_prefix	First 5 digits of Zip Code
geolocation_lat	Latitude
geolocation_lng	Longitude
geolocation_city	City
geolocation_state	State

The **payments.csv** contain following features:

Features	**Description**
order_id	A Unique ID of order made by the consumers
payment_sequential	Sequences of the payments made in case of EMI
payment_type	Mode of payment used (Eg. Credit Card)
payment_installments	Number of installments in case of EMI purchase
payment_value	Total amount paid for the purchase order

The **orders.csv** contain following features:

Features	Description
order_id	A Unique ID of order made by the consumers
customer_id	ID of the consumer who made the purchase
order_status	Status of the order made i.e. delivered, shipped, etc.
order_purchase_timestamp	Timestamp of the purchase
order_delivered_carrier_date	Delivery date at which carrier made the delivery
order_delivered_customer_date	Date at which customer got the product
order_estimated_delivery_date	Estimated delivery date of the products

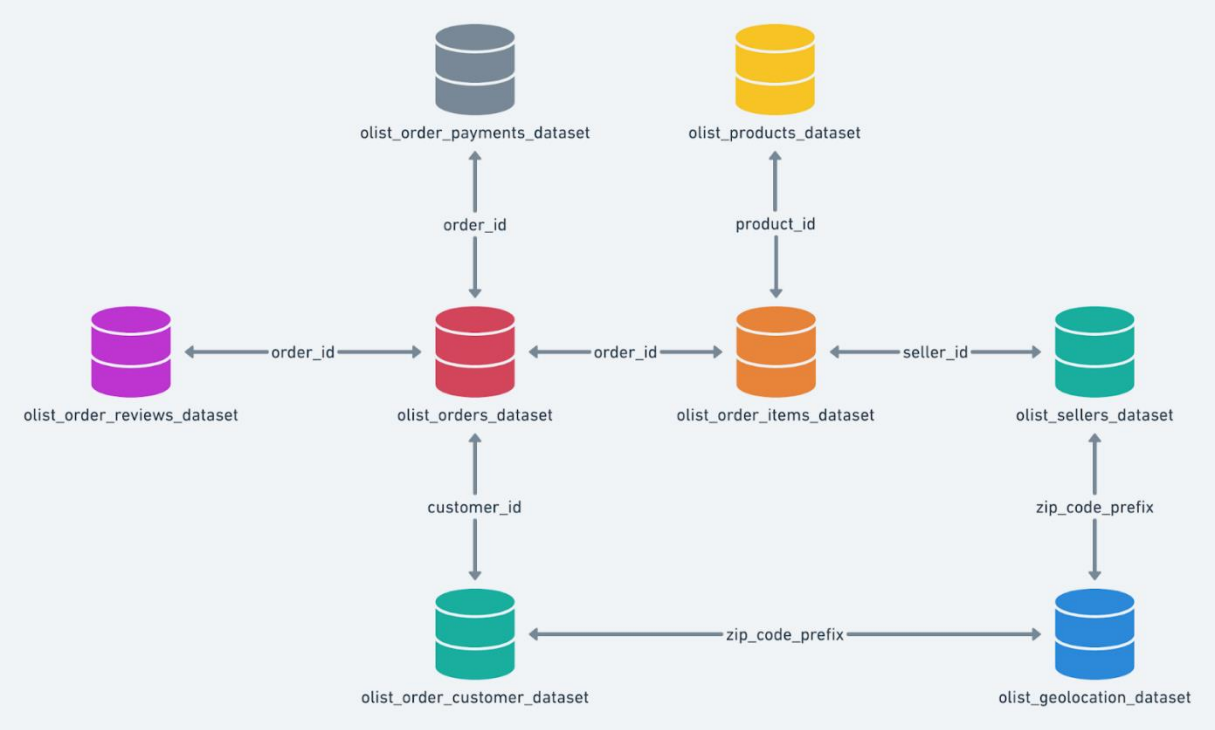
The **reviews.csv** contain following features:

Features	Description
review_id	ID of the review given on the product ordered by the order id
order_id	A Unique ID of order made by the consumers
review_score	Review score given by the customer for each order on a scale of 1-5
review_comment_title	Title of the review
review_comment_message	Review comments posted by the consumer for each order
review_creation_date	Timestamp of the review when it is created
review_answer_timestamp	Timestamp of the review answered

The **products.csv** contain following features:

Features	Description
product_id	A Unique identifier for the proposed project.
product_category_name	Name of the product category
product_name_lenght	Length of the string which specifies the name given to the products o
product_description_lenght	Length of the description written for each product ordered on the site
product_photos_qty	Number of photos of each product ordered available on the shopping
product_weight_g	Weight of the products ordered in grams
product_length_cm	Length of the products ordered in centimeters
product_height_cm	Height of the products ordered in centimeters
product_width_cm	Width of the product ordered in centimeters

Dataset schema:



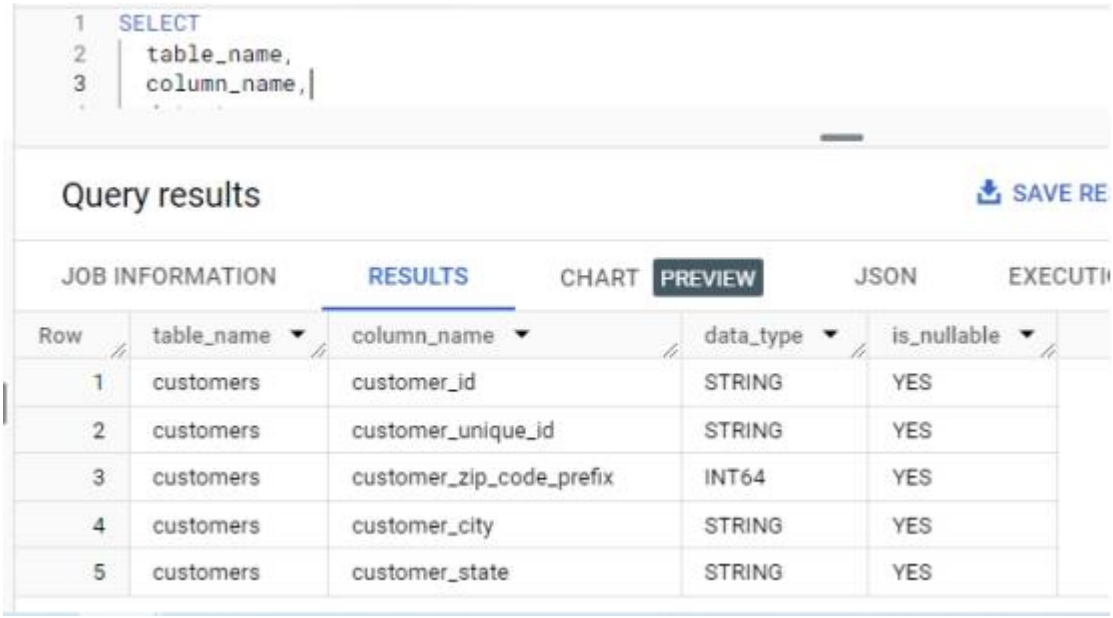
1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:

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

1.1.1. Ans:

```
SELECT
  table_name,
  column_name,
  data_type,
  is_nullable
FROM
  `bigqueryproject-406201.target_sql_business_case.INFORMATION_SCHEMA.COLUMNS`
WHERE
  table_name = 'customers';
```

1.1.1. Screenshots:



Row	table_name	column_name	data_type	is_nullable
1	customers	customer_id	STRING	YES
2	customers	customer_unique_id	STRING	YES
3	customers	customer_zip_code_prefix	INT64	YES
4	customers	customer_city	STRING	YES
5	customers	customer_state	STRING	YES

1.1.2. Insights:

There are 5 columns in customer table.

- customer_zip_code_prefix: INT64
- customer_id, customer_unique_id, customer_city, customer_state: STRING

1.1.3. Recommendations:

1.1.1.1. Input Validation in Application:

- If this data is used in applications, ensure that user inputs for fields like "customer_zip_code_prefix" are validated to be numeric.

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

1.2.1. Ans:

```
select
  min(order_purchase_timestamp) as orders_startDateTime,
  max(order_purchase_timestamp) as orders_endDateTime
from
  `target_sql_business_case.orders`;
```

1.2.2. Screenshots:

```
-- Get the time range between which the orders were placed.
select
  min(order_purchase_timestamp) as orders_startDateTime,
  max(order_purchase_timestamp) as orders_endDateTime
from
  `target_sql_business_case.orders`;
```

ery results

INFORMATION | RESULTS | CHART | PREVIEW | JSON

orders_startDateTime	orders_endDateTime
2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

1.2.3. Insights:

The time span during which the orders were placed extends from September 4, 2016, at 21:15:19 UTC to October 17, 2018, at 17:30:18 UTC.

1.2.4. Recommendations:

1.1.1.2. Temporal Analysis:

- Conduct temporal analysis to identify trends, patterns, or seasonality in order placements over the time.
- Analyze customer behavior changes over time. Identify any shifts or trends in the frequency and timing of orders.

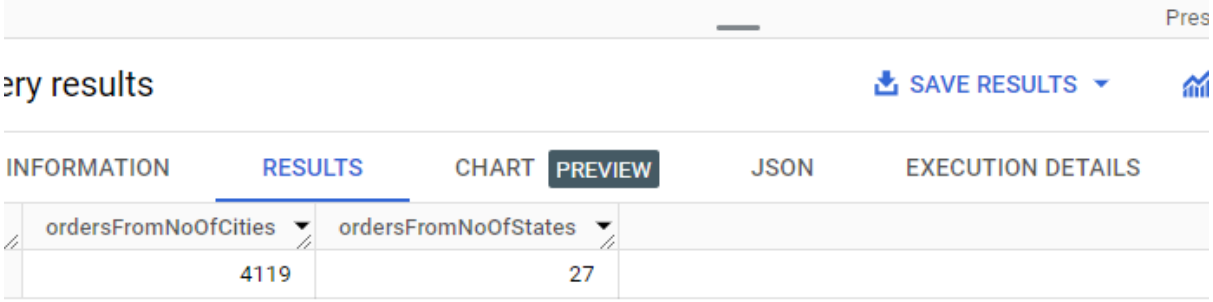
1.3. Count the cities and states of customer who ordered during the given period.

1.3.1. Ans:

```
select
  count(distinct customer_city) ordersFromNoOfCities,
  count(distinct customer_state) ordersFromNoOfStates
from
  `target_sql_business_case.orders` o
  inner join `target_sql_business_case.customers` c on o.customer_id = c.customer_id;
```

1.3.2. Screenshots:

```
--  1.3 Count the Cities & States of customers who ordered during the given period.
select
  count(distinct customer_city) ordersFromNoOfCities,
  count(distinct customer_state) ordersFromNoOfStates
from
  `target_sql_business_case.orders` o
  inner join `target_sql_business_case.customers` c on o.customer_id = c.customer_id;
```



1.3.3. Insights:

We have orders from 4119 cities and 27 states.

1.3.4. Recommendations:

1.1.1.1. Geographic Targeting:

- Leverage the information on cities and states to implement targeted marketing campaigns. Tailor promotions or offers based on the specific preferences or characteristics of customers in different locations.

1.1.1.1. Regional Trends Analysis:

- Analyze order patterns across different states to identify regional trends or preferences. This can help in optimizing inventory, pricing strategies, and product offerings for specific regions.

2. In-depth Exploration:

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

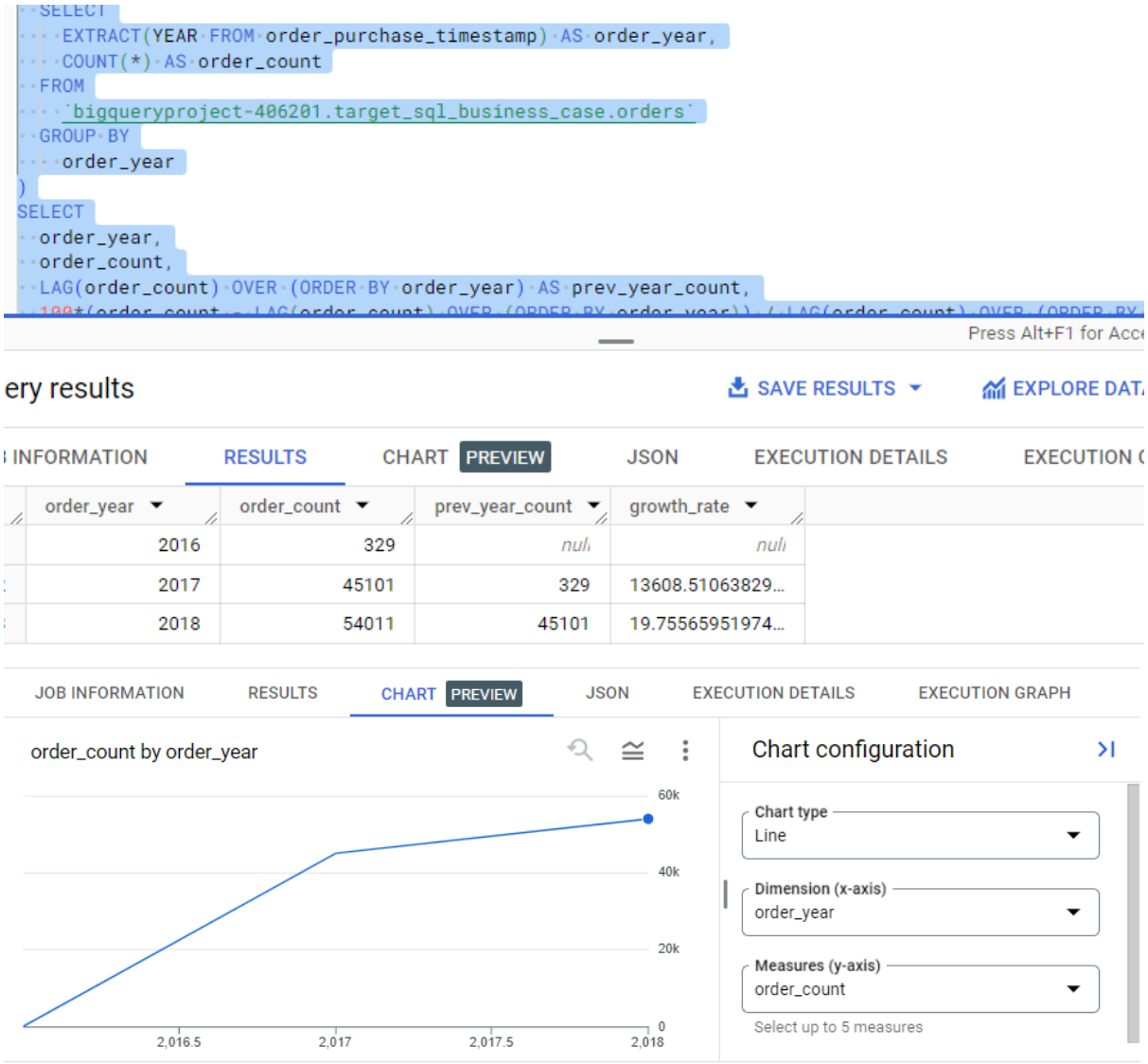
Inshort: Yes, there is a growing trend in the no. of orders placed over the past years. But there are two factors Market capture growth(13,608.51%.) and after market capture growth(19.76%).

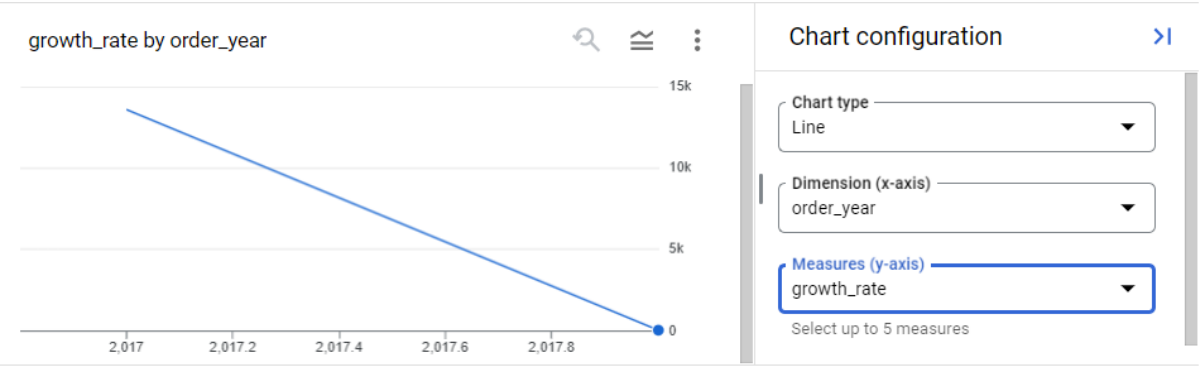
2.1.1. Ans:

```
WITH OrderCounts AS (
  SELECT
    EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
    COUNT(*) AS order_count
  FROM
    `bigqueryproject-406201.target_sql_business_case.orders`
  GROUP BY
    order_year
```

```
)
SELECT
  order_year,
  order_count,
  LAG(order_count) OVER (ORDER BY order_year) AS prev_year_count,
  100*(order_count - LAG(order_count) OVER (ORDER BY order_year)) / LAG(order_count) OVER
(ORDER BY order_year) AS growth_rate
FROM
  OrderCounts
ORDER BY
  order_year;
```

2.1.2. Screenshots:





2.1.3. Insights:

- In 2017, there was a substantial increase in the number of orders compared to 2016, with a growth rate of approximately 13,608.51%.
- In 2018, there was continued growth in the number of orders compared to 2017, but at a lower rate (approximately 19.76%).
- The growth rate indicates fluctuations in order volume year-over-year.

2.1.4. Recommendations:

- Investigate and identify the factors contributing to this significant growth in 2017. Were there specific marketing initiatives, promotions, or improvements in service that drove this surge? Understanding the drivers of success can help in replicating effective strategies.
- Assess the sustainability of this growth rate. Evaluate whether the growth rate is within acceptable ranges and aligns with business goals. Consider diversifying strategies or introducing new initiatives to maintain or accelerate growth.
- Assess the stability of the business. If the growth rate is too erratic, it may be worth considering strategies to smooth out fluctuations. Conversely, if the growth rate is consistently high, strategies can be developed to maintain momentum.

2.1.5. Hypothesis:

Observation: The growth rate in the number of orders was exceptionally high in 2017 (approx. 13,608.51%).

Interpretation: This substantial growth could be indicative of the business capturing a significant market share, entering new markets, implementing successful marketing campaigns, or experiencing a surge in customer acquisition during that period.

2.1.6. Supporting Evidence:

Comparative Analysis: Compare the growth rate in 2017 with industry benchmarks, market trends, or competitor performance to assess if the observed growth is consistent with broader market dynamics.

2.1.7. Market Saturation Consideration:

Analysis: The lower growth rate in 2018 (19.76%) suggests a more stabilized or mature market compared to the rapid expansion in 2017.

Interpretation: The market may have reached a level of saturation, leading to a more normalized growth rate in 2018.

2.1.8. Strategic Implications:

Recommendation: Based on this interpretation, consider adjusting the business strategy. If the market capture was a one-time event, focus on strategies to sustain and retain the captured market share. If the market is maturing, explore diversification strategies, product innovation, or entering new geographic regions.

2.1.9. Customer Retention vs. Acquisition:

Strategy: Assess whether the growth in 2017 was primarily driven by customer acquisition or by retaining existing customers.

Recommendation: Tailor strategies for customer retention and acquisition accordingly. If the growth was driven by acquisitions, consider loyalty programs, personalized marketing, or improved customer service to retain customers.

2.1.10. Forecasting and Planning:

Analysis: The historical data provides insights into past performance.

Recommendation: Use this information to inform future forecasting and strategic planning. Consider whether future growth is expected to be incremental or if there are opportunities for additional surges in specific market conditions.

2.1.11. Competitive Landscape:

Assessment: Understand the competitive landscape during the periods of growth.

Recommendation: If the growth was due to a competitive advantage, assess whether that advantage is sustainable. Keep an eye on competitors and market dynamics that could impact future growth.

It's essential to validate this hypothesis by considering additional contextual information, market research, and a deeper understanding of the business environment. Continuously monitor trends to adapt strategies accordingly.

2.2. Can we see monthly seasonality in terms of the no. of orders being placed?

Inshort: Yes, there is monthly seasonality in terms of the no. of orders being placed. I've noted higher seasonal indices for months like March, May, and July, while lower indices are observed in months like April and June.

Final

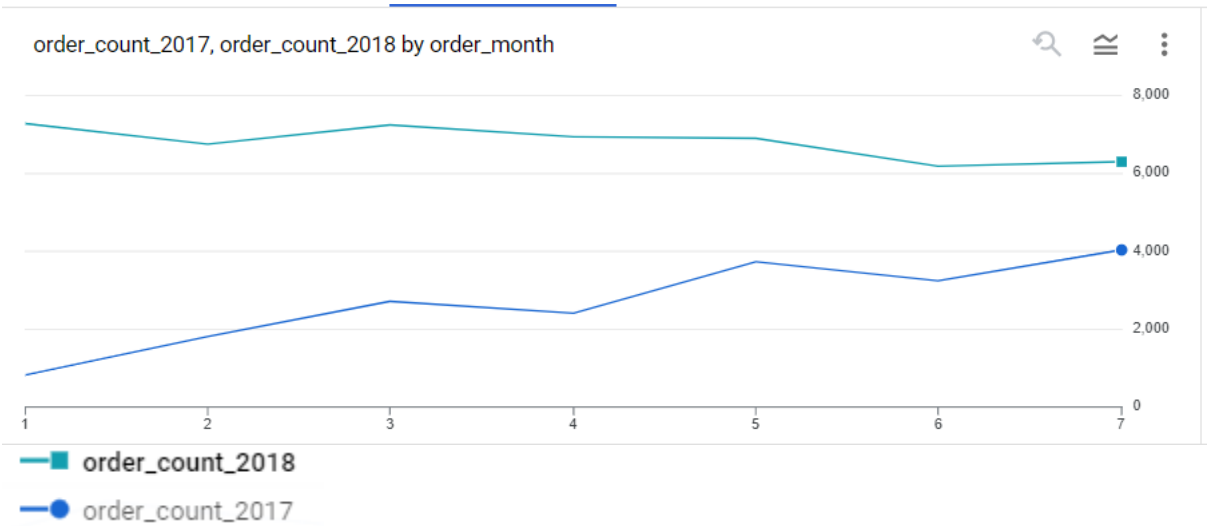
2.2.1. Ans:

```
WITH MonthlyOrderCounts AS (
```

```
  SELECT
    EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
    EXTRACT(month FROM order_purchase_timestamp) AS order_month,
    COUNT(*) AS order_count
  FROM
    `Target.orders`
  GROUP BY
    order_year, order_month
  ORDER BY
    order_year, order_month
)

SELECT
  order_month,
  MAX(CASE WHEN order_year = 2017 THEN order_count END) AS order_count_2017,
  MAX(CASE WHEN order_year = 2018 THEN order_count END) AS order_count_2018
FROM
  MonthlyOrderCounts
WHERE
  order_month BETWEEN 1 AND 7
GROUP BY
  order_month
ORDER BY
  order_month;
```

2.2.2. Screenshots:

Row	order_month	order_count_2017	order_count_2018
1	1	800	7269
2	2	1780	6728
3	3	2682	7211
4	4	2404	6939
5	5	3700	6873
6	6	3245	6167
7	7	4026	6292



Analysis of the filtered data for the months from January to July reveals a pattern of monthly seasonality. Specifically, I've noted higher seasonal indices for months like March, May, and July, while lower indices are observed in months like April and June. This pattern suggests that there might be certain months where Brazilian customers are more active in placing orders.

Here are some insights and recommendations based on findings:

2.2.2.1. Identified Monthly Seasonality:

- **Insight**: March, May, and July show higher order counts compared to the surrounding months, indicating a potential seasonal pattern.
- **Recommendation**: Explore the reasons behind the increased activity in these months. Consider aligning marketing campaigns, promotions, or special events to leverage this observed seasonality.

2.2.2.2. Lower Indices in April and June:

- **Insight**: April and June exhibit lower order counts based on the seasonal indices.
- **Recommendation**: Investigate factors contributing to the lower activity in these months. Consider implementing targeted marketing strategies or promotions to boost order counts during these periods.

2.2.2.3. Fine-Tune Strategies:

- **Insight**: Monthly seasonality implies that customer behaviour varies throughout the year.
- **Recommendation**: Fine-tune your marketing and operational strategies to align with the observed monthly patterns. This might involve adjusting advertising spend, inventory management, or customer engagement initiatives.

2.2.2.4. Promotions and Campaigns:

- **Insight**: Higher order counts in specific months suggest opportunities for targeted promotions or campaigns.
- **Recommendation**: Plan promotions, discounts, or special events around the months with higher customer activity. Consider creating incentives to drive additional sales during months with traditionally lower order counts.

2.2.2.5. Customer Engagement:

- **Insight**: Understanding seasonal trends can help optimize customer engagement efforts.
- **Recommendation**: Tailor communication and engagement strategies based on the observed monthly patterns. This could include personalized marketing messages or loyalty programs designed to align with customer behaviour.

2.2.2.6. Continuous Monitoring:

- **Insight**: Monthly seasonality may evolve over time.
- **Recommendation**: Continuously monitor order patterns and adjust strategies accordingly. Regularly analyse data to identify changes in customer behaviour or external factors influencing order counts.

My steps to reach the final

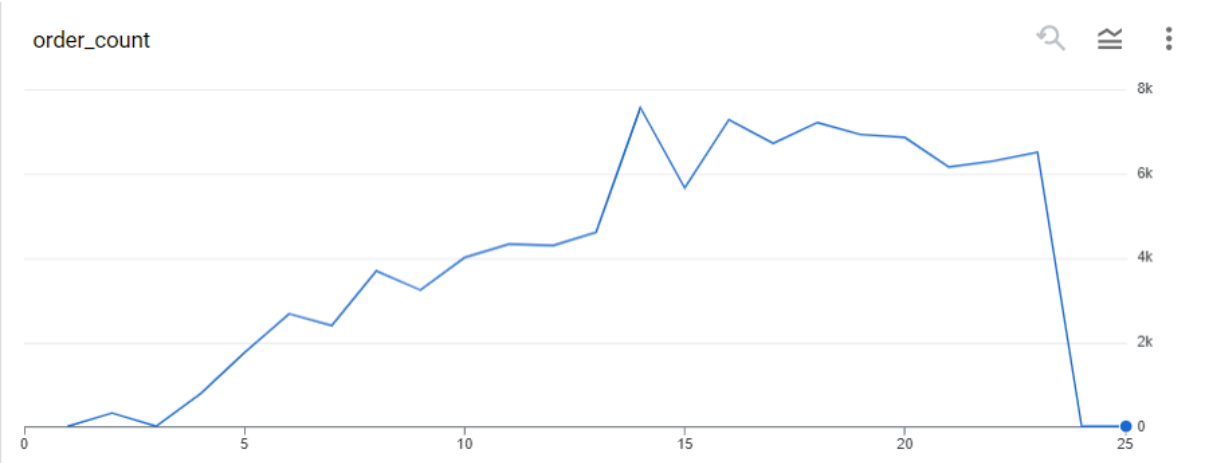
Part-1

2.2.3. Ans:

```
SELECT
    EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
    EXTRACT(month FROM order_purchase_timestamp) AS order_month,
    COUNT(*) AS order_count
  FROM
    `bigqueryproject-406201.target_sql_business_case.orders`
  GROUP BY
    order_year, order_month
  ORDER BY
    1,2
```

2.2.4. Screenshots:

Row	order_year	order_month	order_count
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



2.2.5. Insights:

2.2.5.1. Monthly Order Counts:

- In 2016, the order counts are relatively low, with a spike in October (324 orders).
- In 2017, there is a general upward trend in order counts throughout the year, reaching a peak in November (7544 orders).
- In 2018, the trend continues with high order counts until March, after which there is a noticeable decline.

2.2.5.2. Key Observations:

Sharp Spike in November 2017:

- A substantial increase in orders is observed in November 2017, reaching 7544 orders. This could be attributed to various factors such as Black Friday, Cyber Monday, or holiday season promotions.

Balancing Out in December 2017:

- In December 2017, the order count balances out, dropping to 5673 orders. This could be a result of seasonal adjustments or the normalization of order patterns after a peak month.

Downward Trend Starting in March 2018:

- After reaching a peak in March 2018, there is a noticeable downward trend in order counts. This downward trend continues, with a significant drop in September 2018 (16 orders) and an even lower count in October 2018 (4 orders).

2.2.6. Recommendations:

2.2.6.1. Seasonal Adjustments:

- Consider adjusting strategies for seasonal fluctuations, especially during months with historically high order counts. Plan promotions, marketing campaigns, and inventory accordingly.

2.2.6.2. Investigate the November Spike:

- Investigate the factors contributing to the sharp spike in November 2017. Identify successful promotions or external factors that drove this surge to replicate effective strategies.

2.2.6.3. Addressing Downward Trend:

- Analyze the reasons for the downward trend starting in March 2018 and the significant drop in September 2018. Identify potential issues affecting customer engagement or external factors impacting demand.

2.2.6.4. Customer Retention:

- Consider strategies for customer retention, especially during periods of declining order counts. Implement loyalty programs, personalized marketing, or targeted promotions to re-engage existing customers.

2.2.6.5. Forecasting and Planning:

- Use historical data to refine forecasting models. Understand the patterns and trends to anticipate future order volumes more accurately.

Part-2

2.2.7. Ans:

```
WITH MonthlyOrderCounts AS (
  SELECT
    EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
    EXTRACT(month FROM order_purchase_timestamp) AS order_month,
    COUNT(*) AS order_count
  FROM
    `bigqueryproject-406201.target_sql_business_case.orders`
  GROUP BY
    order_year, order_month
  ORDER BY
    1,2
)
```

```sql
SELECT
  order_year,
  order_month,
  order_count,
  ROUND(AVG(order_count) OVER (), 2) AS average_order_count,
  ROUND((order_count / AVG(order_count) OVER ()) * 100, 2) AS seasonal_index
FROM
  MonthlyOrderCounts
ORDER BY
  order_month, order_year

;
```

2.2.8. Screenshots:

Row	order_year	order_month	order_count	average_order_count	seasonal_index
1	2017	1	800	4505.09	17.76
2	2018	1	7269	4505.09	161.35
3	2017	2	1780	4505.09	39.51
4	2018	2	6728	4505.09	149.34
5	2017	3	2682	4505.09	59.53
6	2018	3	7211	4505.09	160.06
7	2017	4	2404	4505.09	53.36
8	2018	4	6939	4505.09	154.03
9	2017	5	3700	4505.09	82.13

2.2.9. Insights:

- **seasonal_index** : A value greater than 100 indicates above-average activity, while a value below 100 indicates below-average activity.
- In this part we can see the seasonal_index but unable to compare it lets do it in next part.

2.2.10. Recommendations:

- Get the index_difference on this data to get some pattern.

Part-3

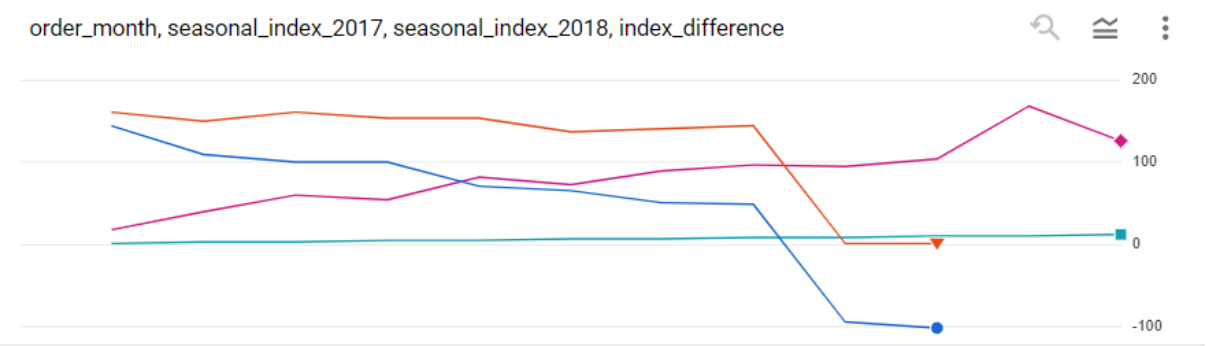
2.2.11. Ans:

```sql
WITH MonthlyOrderCounts AS (
  SELECT
    EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
    EXTRACT(month FROM order_purchase_timestamp) AS order_month,
    COUNT(*) AS order_count
  FROM
    `bigqueryproject-406201.target_sql_business_case.orders`
  GROUP BY
    order_year, order_month
  ORDER BY
    1,2
```

```
),

seasonal_index_table AS(
    SELECT
      order_year,
      order_month,
      order_count,
      ROUND(AVG(order_count) OVER (), 2) AS average_order_count,
      ROUND((order_count / AVG(order_count) OVER ()) * 100, 2) AS seasonal_index
    FROM
      MonthlyOrderCounts
    WHERE order_year IN (2017, 2018)
    ORDER BY
      order_month,order_year
)

SELECT
  order_month,
  AVG(CASE WHEN order_year = 2017 THEN seasonal_index END) AS seasonal_index_2017,
  AVG(CASE WHEN order_year = 2018 THEN seasonal_index END) AS seasonal_index_2018,
  ROUND(AVG(CASE WHEN order_year = 2018 THEN seasonal_index END) -
        AVG(CASE WHEN order_year = 2017 THEN seasonal_index END), 2) AS index_difference
FROM seasonal_index_table
GROUP BY
  order_month
ORDER BY
  order_month;
```

2.2.12. Screenshots:

Row	order_month	seasonal_index_2017	seasonal_index_2018	index_difference
1	1	17.76	161.35	143.59
2	2	39.51	149.34	109.83
3	3	59.53	160.06	100.53
4	4	53.36	154.03	100.67
5	5	82.13	152.56	70.43
6	6	72.03	136.89	64.86
7	7	89.37	139.66	50.29
8	8	96.14	144.55	48.41



2.2.13. Monthly Seasonality Absence:

- **Insight**: The comparison of seasonal indices between 2017 and 2018 reveals that there is no clear monthly seasonality in terms of the number of orders being placed.
- **Recommendation**: Given the absence of clear monthly seasonality, focus on understanding the factors contributing to the overall trend and explore strategies to address any challenges affecting order counts.

2.2.14. Downtrend Analysis:

- **Insight**: The downtrend observed in the index difference is mainly driven by a sharp drop in the seasonal index in September 2018.
- **Recommendation**: Investigate the specific factors contributing to the drop in September 2018. This could include external factors, changes in consumer behaviour, or internal operational challenges. Identifying the root cause will be crucial for implementing targeted improvements.

2.2.15. September 2018 Drop:

- **Insight**: September 2018 has a significant negative index difference, indicating a substantial decline in the seasonal index compared to the same month in 2017.
- **Recommendation**: Conduct a detailed analysis of the events, promotions, or operational aspects that occurred in September 2018. Understand whether there were any external factors impacting customer engagement during that period.

2.2.16. Strategic Adjustments:

- **Insight**: The negative index differences in September and October 2018 suggest a challenging period during those months.
- **Recommendation**: Based on the insights, consider implementing strategic adjustments. This could involve revisiting marketing strategies, promotional activities, or customer engagement initiatives to counter the observed decline.

2.2.17. Future Monitoring:

- **Insight**: The trend analysis highlights the importance of monitoring and adapting strategies based on changing circumstances.
- **Recommendation**: Establish a robust monitoring system to track key performance indicators (KPIs) and promptly identify shifts in customer behaviour, market dynamics, or internal operations. This proactive approach will enable timely responses to evolving trends.

Part-4

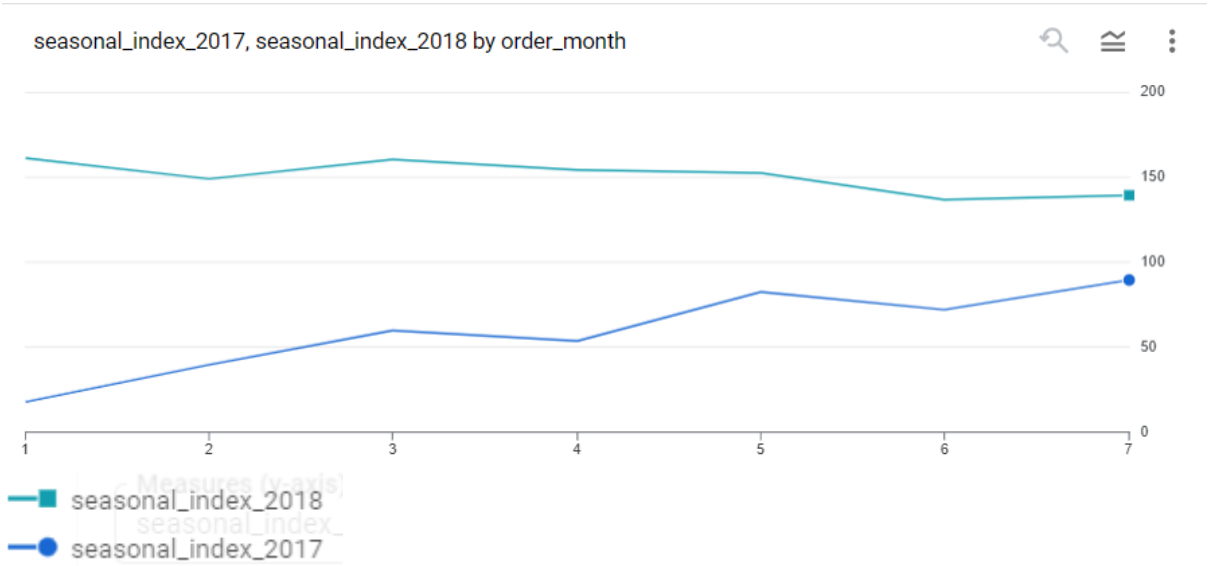
2.2.18. Ans:

```
WITH MonthlyOrderCounts AS (
  SELECT
    EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
    EXTRACT(month FROM order_purchase_timestamp) AS order_month,
    COUNT(*) AS order_count
  FROM
    `bigqueryproject-406201.target_sql_business_case.orders`
  GROUP BY
    order_year, order_month
  ORDER BY
    1,2
),

seasonal_index_table AS(
    SELECT
      order_year,
      order_month,
      order_count,
      ROUND(AVG(order_count) OVER (), 2) AS average_order_count,
      ROUND((order_count / AVG(order_count) OVER ()) * 100, 2) AS seasonal_index
    FROM
      MonthlyOrderCounts
    WHERE order_year IN (2017, 2018)
    ORDER BY
      order_month,order_year
)

SELECT
  order_month,
  AVG(CASE WHEN order_year = 2017 THEN seasonal_index END) AS seasonal_index_2017,
  AVG(CASE WHEN order_year = 2018 THEN seasonal_index END) AS seasonal_index_2018
FROM seasonal_index_table
WHERE order_month BETWEEN 1 AND 7
GROUP BY
```

```
  order_month
ORDER BY
  order_month;
```

2.2.19. Screenshots:

Row	order_month	seasonal_index_201	seasonal_index_201
1	1	17.76	161.35
2	2	39.51	149.34
3	3	59.53	160.06
4	4	53.36	154.03
5	5	82.13	152.56
6	6	72.03	136.89
7	7	89.37	139.66



2.2.20. Insights:

By looking at the filtered data for the month jan to July monthly seasonality can be seen. There are months like march, may and july where indexes are higher and and months like April and June are lower.

2.2.21. Recommendations:

2.3. During what time of the day, do the Brazilian customers mostly place their orders? (0-6 Dawn, 7-12 Morning, 13-18 Afternoon or 19-23 Night)

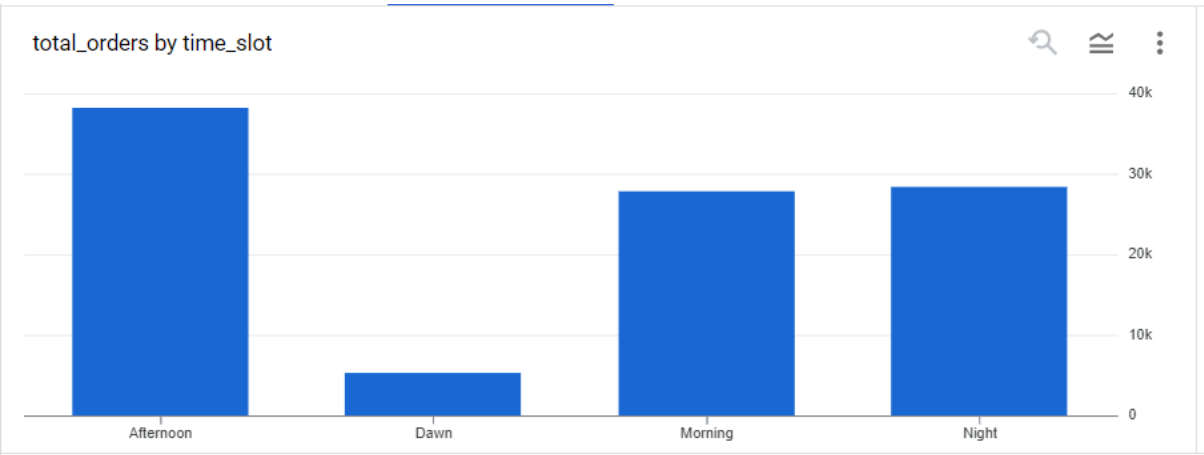
2.3.1. Ans:

```
WITH OrderTimeSlots AS (
  SELECT
    EXTRACT(HOUR FROM order_purchase_timestamp) AS order_hour,
    COUNT(*) AS order_count
  FROM
```

```
    `Target.orders`
  GROUP BY
    order_hour
)

SELECT
  CASE
    WHEN order_hour >= 0 AND order_hour < 7 THEN 'Dawn'
    WHEN order_hour >= 7 AND order_hour < 13 THEN 'Morning'
    WHEN order_hour >= 13 AND order_hour < 19 THEN 'Afternoon'
    WHEN order_hour >= 19 AND order_hour <= 23 THEN 'Night'
    ELSE 'Unknown'
  END AS time_slot,
  SUM(order_count) AS total_orders
FROM
  OrderTimeSlots
GROUP BY
  time_slot
ORDER BY
  time_slot;
```

2.3.2. Screenshots:

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



2.3.3. Peak Order Placement in the Afternoon:

Insight: The highest total number of orders occurs in the afternoon.

Recommendation: Leverage this information to optimize resource allocation, such as customer support staffing, order fulfilment processes, and inventory management during peak hours.

2.3.4. Understanding Customer Behaviour:

Insight: The distribution of orders across different time slots reflects customer behaviour.

Recommendation: Consider conducting further analysis to understand why customers prefer certain times of the day for making purchases. This insight can guide personalized marketing strategies or promotions targeted at specific times.

2.3.5. Marketing and Promotions:

Insight: Afternoon sees the highest order activity.

Recommendation: Schedule and focus marketing campaigns, promotions, and product launches during the afternoon to maximize visibility and engagement when customers are most active.

2.3.6. Optimizing Delivery and Logistics:

Insight: Knowing peak order times helps optimize delivery and logistics.

Recommendation: Ensure that delivery and fulfilment services are well-prepared to handle the higher order volumes during the afternoon. Consider offering expedited or same-day delivery options during peak hours.

2.3.7. Customer Support Availability:

Insight: Afternoon is a critical time for order placement.

Recommendation: Ensure that customer support is adequately staffed during the afternoon to address inquiries, provide assistance, and enhance the overall customer experience.

2.3.8. Continuous Monitoring:

Insight: Customer behaviour may evolve over time.

Recommendation: Continuously monitor order patterns and analyse trends. Be prepared to adapt strategies based on changes in customer behaviour or external factors influencing order placement.

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

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

3.1.1. Ans:

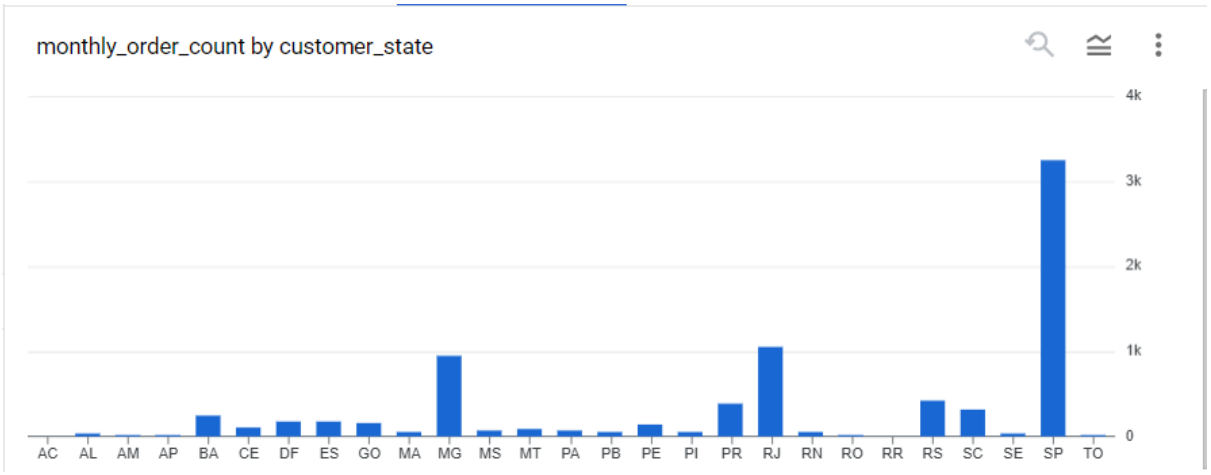
```
WITH MonthlyOrderCounts AS (
  SELECT
    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
    c.customer_state,
    COUNT(*) AS order_count
  FROM
    `Target.orders` o
    JOIN `Target.customers` c ON o.customer_id = c.customer_id
  GROUP BY
    customer_state, order_year, order_month
  ORDER BY
    customer_state, order_year, order_month
```

```
)

SELECT
  customer_state,
  order_year,
  order_month,
  order_count AS monthly_order_count,
  SUM(order_count) OVER (PARTITION BY customer_state ORDER BY order_year, order_month) AS
cumulative_orders
FROM
  MonthlyOrderCounts
ORDER BY
  customer_state, order_year, order_month;
```

3.1.2. Screenshots:

Row	customer_state	order_year	order_month	monthly_order_count	cumulative_orders
1	AC	2017	1	2	2
2	AC	2017	2	3	5
3	AC	2017	3	2	7
4	AC	2017	4	5	12
5	AC	2017	5	8	20
6	AC	2017	6	4	24
7	AC	2017	7	5	29
8	AC	2017	8	4	33
9	AC	2017	9	5	38
10	AC	2017	10	6	44

Row	customer_state	order_year	order_month	monthly_order_count	cumulative_orders
16	AC	2018	4	4	69
17	AC	2018	5	2	71
18	AC	2018	6	3	74
19	AC	2018	7	4	78
20	AC	2018	8	3	81
21	AL	2016	10	2	2
22	AL	2017	1	2	4
23	AL	2017	2	12	16
24	AL	2017	3	10	26



3.1.3. Insights:

Regional Disparities: There are clear regional variations in e-commerce activity. São Paulo (SP) dominates in terms of monthly order counts, while states like Minas Gerais (MG), Paraná (PR), Rio de Janeiro (RJ), Rio Grande do Sul (RS), and Santa Catarina (SC) also show substantial order volumes.

Opportunities for Growth: States with lower order counts represent untapped markets. Exploring marketing strategies, promotions, or partnerships specific to these regions could unlock growth potential.

Customer Engagement: Understanding the factors contributing to high order counts in leading states can provide insights into customer behavior. Analyzing popular products, marketing channels, and customer demographics can help tailor strategies to maximize engagement.

Seasonal Trends: Examining monthly patterns reveals potential seasonal influences. Adjusting inventory, marketing, and promotional efforts to align with these trends can enhance efficiency.

3.1.4. Recommendations:

Targeted Marketing: Develop targeted marketing campaigns for states with lower order counts. Consider region-specific promotions, partnerships with local businesses, or culturally relevant advertisements to attract more customers.

Product Assortment: Assess the popularity of products in different regions. Offering region-specific or culturally relevant products might resonate better with customers and boost sales.

Customer Outreach: Implement strategies to engage customers in states with lower order counts. This could include personalized promotions, loyalty programs, or surveys to understand customer needs and preferences.

Logistics Optimization: Evaluate and optimize logistics and delivery services in high-order states to maintain efficiency. This ensures a positive customer experience and encourages repeat business.

Market Expansion: Explore opportunities for market expansion in states with high order counts. This could involve opening new distribution centers, expanding product offerings, or strengthening partnerships with local businesses.

Data-Driven Decision Making: Continue using data analytics to inform decision-making. Regularly review and update strategies based on changing customer behaviors, market trends, and competition.

Extra

To visualize the evolution of e-commerce orders for each state and compare them, we can use various data visualization tools or libraries. Popular tools include Matplotlib (for Python), Tableau, or Google Data Studio. Still I have plotted the grapes for State SP.

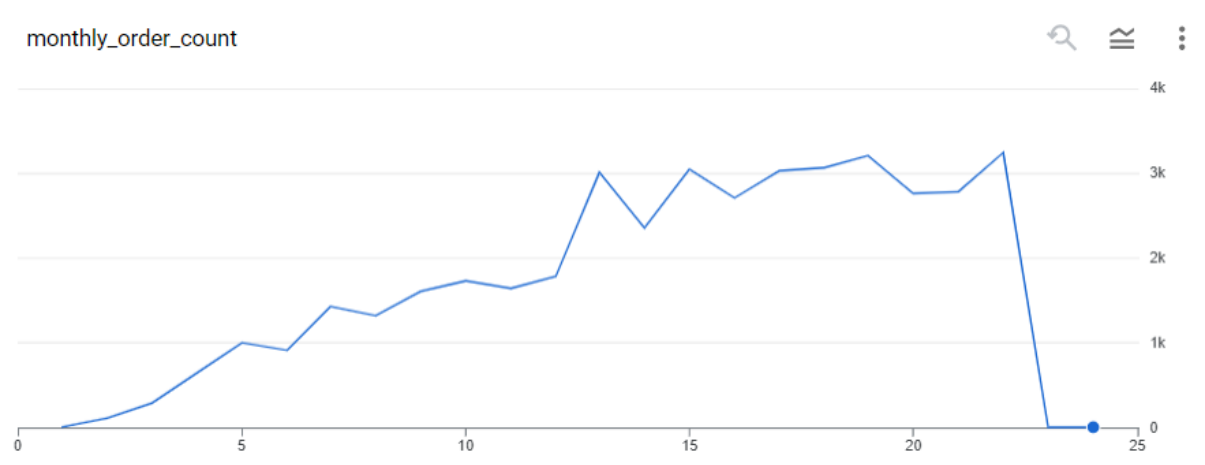
3.1.5. Ans:

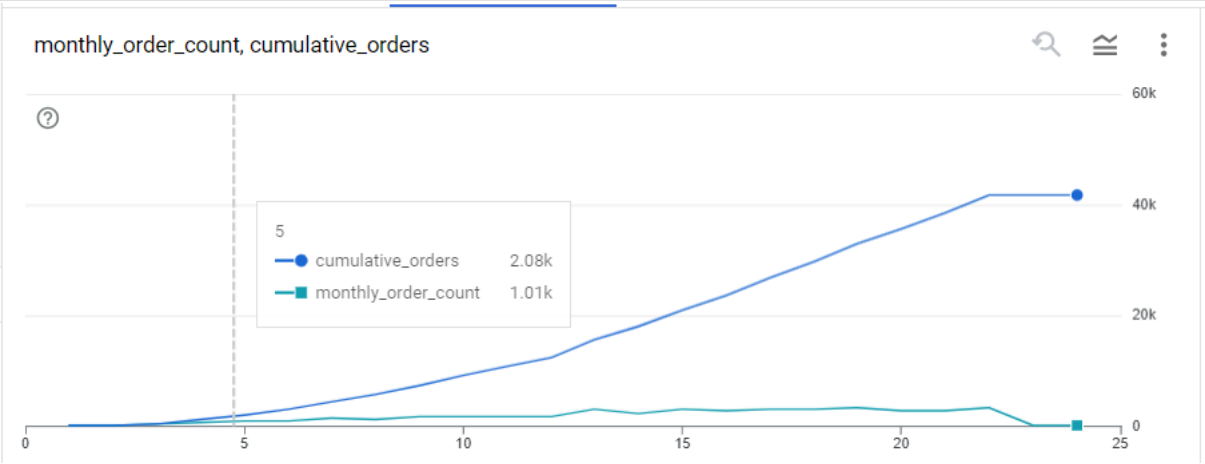
```
WITH MonthlyOrderCounts AS (
  SELECT
    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
    c.customer_state,
    COUNT(*) AS order_count
  FROM
    `Target.orders` o
    JOIN `Target.customers` c ON o.customer_id = c.customer_id
  GROUP BY
    customer_state, order_year, order_month
  ORDER BY
    customer_state, order_year, order_month
)

SELECT
  customer_state,
  order_year,
  order_month,
  order_count AS monthly_order_count,
  SUM(order_count) OVER (PARTITION BY customer_state ORDER BY order_year, order_month) AS
cumulative_orders
FROM
  MonthlyOrderCounts
```

```
WHERE customer_state = 'SP'
ORDER BY
  customer_state, order_year, order_month;
```

3.1.6. Screenshots:

Row	customer_state	order_year	order_month	monthly_order_count	cumulative_orders
1	SP	2016	9	2	2
2	SP	2016	10	113	115
3	SP	2017	1	299	414
4	SP	2017	2	654	1068
5	SP	2017	3	1010	2078
6	SP	2017	4	908	2986
7	SP	2017	5	1425	4411
8	SP	2017	6	1331	5742
9	SP	2017	7	1604	7346
10	SP	2017	8	1729	9075





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

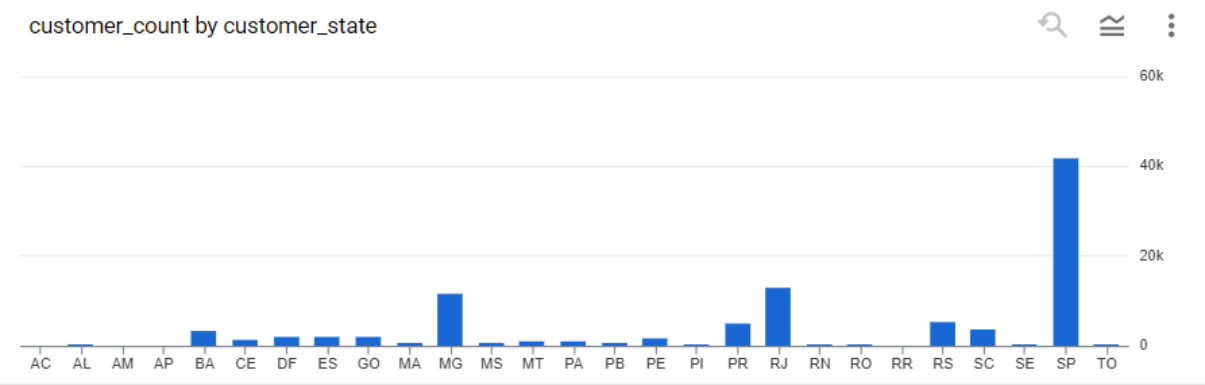
3.2.1. Ans:

```
SELECT
  customer_state,
  COUNT(DISTINCT customer_id) AS customer_count
FROM
  `Target.customers`
GROUP BY
  customer_state
```

```
ORDER BY
  customer_state;
```

3.2.2. Screenshots:

Row	customer_state	customer_count
1	AC	81
2	AL	413
3	AM	148
4	AP	68
5	BA	3380
6	CE	1336
7	DF	2140
8	ES	2033
9	GO	2020
10	MA	747



3.2.3. Customer Concentration in SP:

Insight: São Paulo (SP) has the highest number of customers by a significant margin.

Recommendation: Leverage the large customer base in São Paulo for targeted marketing campaigns, promotions, and product launches. Consider tailoring strategies to meet the preferences and needs of this substantial customer segment.

3.2.4. Significant Customer Base in MG, PR, RJ, and RS:

Insight: States like Minas Gerais (MG), Paraná (PR), Rio de Janeiro (RJ), and Rio Grande do Sul (RS) also have substantial customer counts.

Recommendation: Focus on maintaining and expanding the customer base in these states. Explore strategies to enhance customer loyalty, gather feedback, and understand regional preferences.

3.2.5. Opportunities in Other States:

Insight: Some states have lower customer counts.

Recommendation: Identify opportunities to attract and engage customers in states with lower counts. Consider targeted marketing efforts, localized promotions, or region-specific initiatives to increase brand awareness and customer acquisition.

3.2.6. Segmented Marketing Strategies:

Insight: The customer distribution varies widely across states.

Recommendation: Develop segmented marketing strategies that take into account the regional differences. Tailor promotions, discounts, and communication to align with the characteristics and preferences of customers in each state.

3.2.7. Monitoring and Adaptation:

Insight: Customer distribution provides a snapshot of the current state.

Recommendation: Continuously monitor customer distribution trends over time. Be prepared to adapt strategies based on changes in customer behavior, market dynamics, or external factors influencing customer acquisition.

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 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.

4.1.1. Ans:

```
WITH OrderCosts AS (
  SELECT
    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
    p.payment_value
  FROM
    `Target.orders` o
    JOIN `Target.payments` p ON o.order_id = p.order_id
  WHERE
    EXTRACT(YEAR FROM o.order_purchase_timestamp) IN (2017, 2018)
    AND EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
)

SELECT
  order_year,
  ROUND(SUM(payment_value), 2) AS cost,
  LAG(SUM(payment_value), 1) OVER (ORDER BY order_year) AS lag_cost,
  ROUND(((SUM(payment_value) - LAG(SUM(payment_value), 1) OVER (ORDER BY order_year)) /
LAG(SUM(payment_value), 1) OVER (ORDER BY order_year)) * 100, 2) AS percentage_increase
FROM
  OrderCosts
GROUP BY
  order_year
```

```
ORDER BY
  order_year;
```

4.1.2. Screenshots:

Row	order_year	cost	lag_cost	percentage_increase
1	2017	3669022.12	null	null
2	2018	8694733.84	3669022.119999...	136.98

4.1.3. Insights:

Percentage Increase: The percentage increase from 2017 to 2018 is approximately 136.98%. This indicates a substantial rise in the overall cost of orders during this period.

Substantial Growth in 2018: The total cost of orders in 2018 experienced a significant increase compared to 2017, with a growth from approximately 3.67 million to 8.69 million units of currency.

4.1.4. Recommendations:

Identify Factors Driving Growth: Investigate the specific factors contributing to the remarkable growth in the cost of orders. This could include increased customer acquisition, higher average order values, or expanded product offerings.

Optimize Marketing and Sales Strategies: If the growth is attributed to successful marketing or sales strategies, consider optimizing and expanding these approaches. Identify which channels, promotions, or campaigns contributed most to the growth and allocate resources accordingly.

Customer Engagement and Retention: Assess the impact of customer engagement and retention efforts on the growth. Implement strategies to enhance customer loyalty, satisfaction, and repeat business.

Product Portfolio Management: Evaluate the performance of different product categories. If certain categories contributed significantly to the growth, focus on optimizing the product portfolio and exploring opportunities for expansion.

Market Trends and Competition: Stay informed about market trends and monitor competitors to ensure that the growth aligns with broader industry dynamics. Adjust strategies based on market conditions and evolving customer preferences.

Forecasting and Planning: Use the insights gained to improve forecasting and planning for future periods. Align inventory, staffing, and other resources with the expected demand to ensure smooth operations.

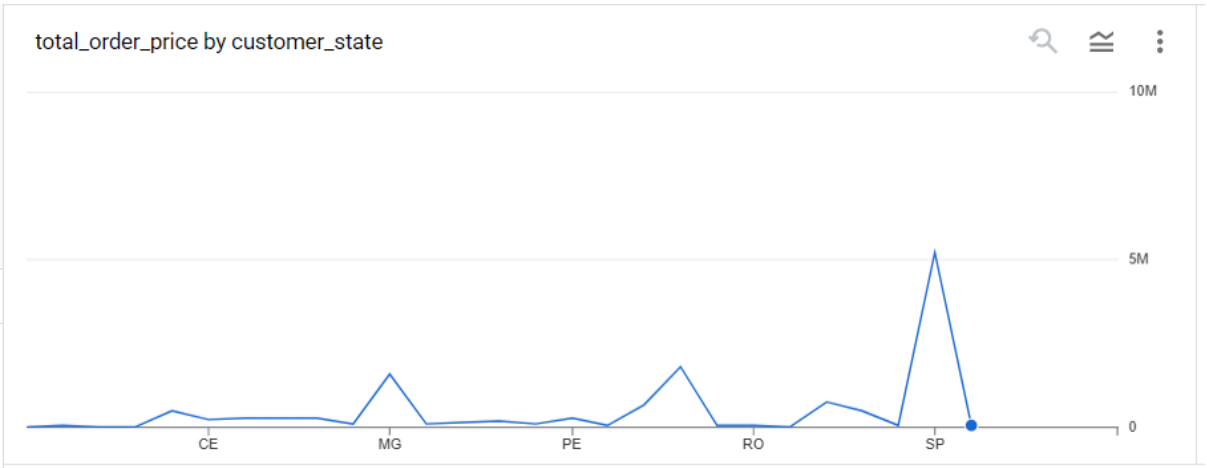
4.2. Calculate the Total & Average value of order price for each state.

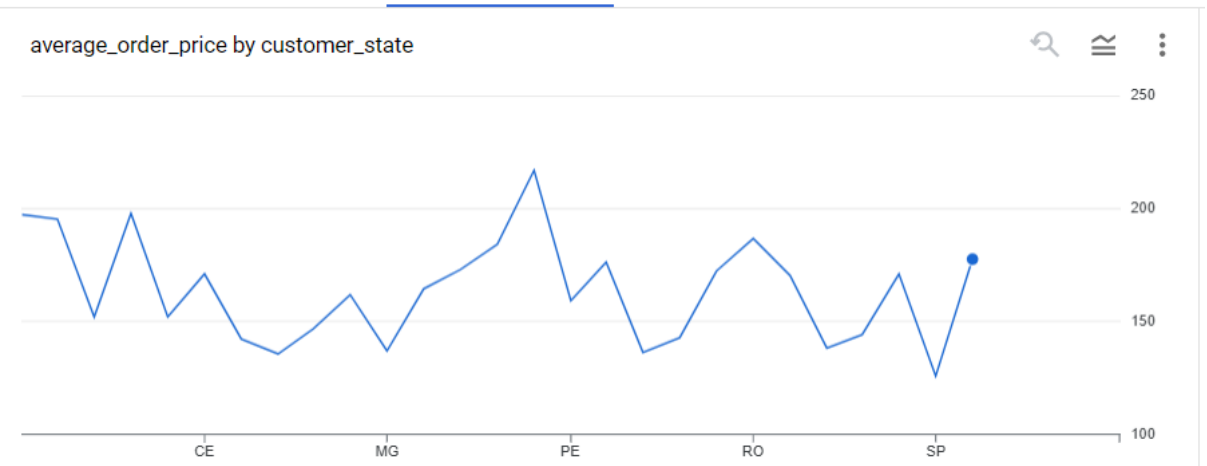
4.2.1. Ans:

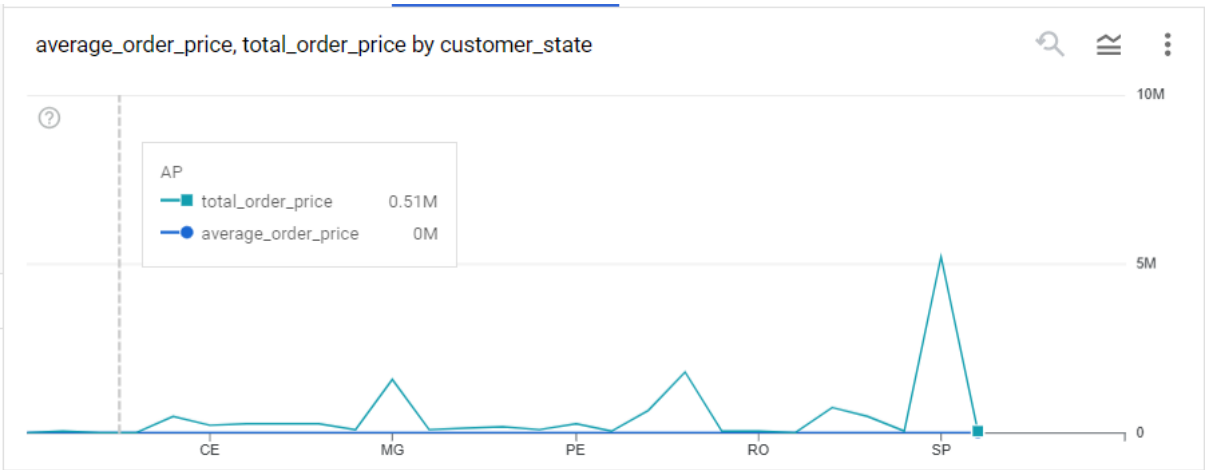
```
WITH OrderPrices AS (
  SELECT
    c.customer_state,
    oi.order_id,
    SUM(oi.price) AS order_total_price
  FROM
    `Target.customers` c
    JOIN `Target.orders` o ON c.customer_id = o.customer_id
    JOIN `Target.order_items` oi ON o.order_id = oi.order_id
  GROUP BY
    c.customer_state, oi.order_id
)

SELECT
  customer_state,
  ROUND(SUM(order_total_price), 2) AS total_order_price,
  ROUND(AVG(order_total_price), 2) AS average_order_price
FROM
  OrderPrices
GROUP BY
  customer_state
ORDER BY
  customer_state;
```

4.2.2. Screenshots:

Row	customer_state	total_order_price	average_order_price
1	AC	15982.95	197.32
2	AL	80314.81	195.41
3	AM	22356.84	152.09
4	AP	13474.3	198.15
5	BA	511349.99	152.28
6	CE	227254.71	171.25
7	DF	302603.94	142.4
8	ES	275037.31	135.82
9	GO	294591.95	146.78
10	MA	119648.22	161.69







4.2.3. Insights:

4.2.3.1. Average Order Price Distribution:

- The average order price is generally between 150 to 200 units of currency across states.
- The state of Paraíba (PB) has the highest average order price, indicating that customers in this state tend to spend more per order.

- States like Amapá (AP), Roraima (RR), Rondônia (RO), and Sergipe (SE) also exhibit higher average order prices.

4.2.3.2. Total Order Price Contribution:

- São Paulo (SP) significantly contributes to the total order price, with the highest total order price among all states.
- Other major contributors include Minas Gerais (MG), Rio de Janeiro (RJ), Bahia (BA), and Rio Grande do Sul (RS).
- States like Acre (AC), Amapá (AP), and Roraima (RR) contribute the least to the total order price.

4.2.4. Recommendations:

4.2.4.1. Focus on High-Average States:

- Consider implementing targeted marketing strategies in states with higher average order prices (e.g., PB, AP, RO) to capitalize on customers willing to spend more.

4.2.4.2. Optimize Pricing in Low-Average States:

- In states with lower average order prices (e.g., ES, MG, PR, RS, SP), explore pricing strategies, discounts, or promotions to encourage higher spending per order.

4.2.4.3. Enhance Marketing Efforts in High-Contributing States:

- States like SP, MG, RJ, BA, and RS contribute significantly to the total order price. Strengthen marketing and promotional activities in these states to maintain and potentially increase their contribution.

4.2.4.4. Customer Segmentation:

- Analyze customer demographics and preferences in high-contributing states to tailor products and services to specific customer segments.

4.2.4.5. Evaluate Product Mix:

- Assess the product mix in each state to understand which products contribute most to the order value. Optimize the product portfolio based on local preferences.

4.2.4.6. Promote Cross-Selling:

- Implement cross-selling strategies to encourage customers to purchase complementary products, potentially increasing the overall order value.

4.2.4.7. Monitor and Adjust:

- Continuously monitor customer behavior, market trends, and competition to adapt pricing and marketing strategies accordingly.

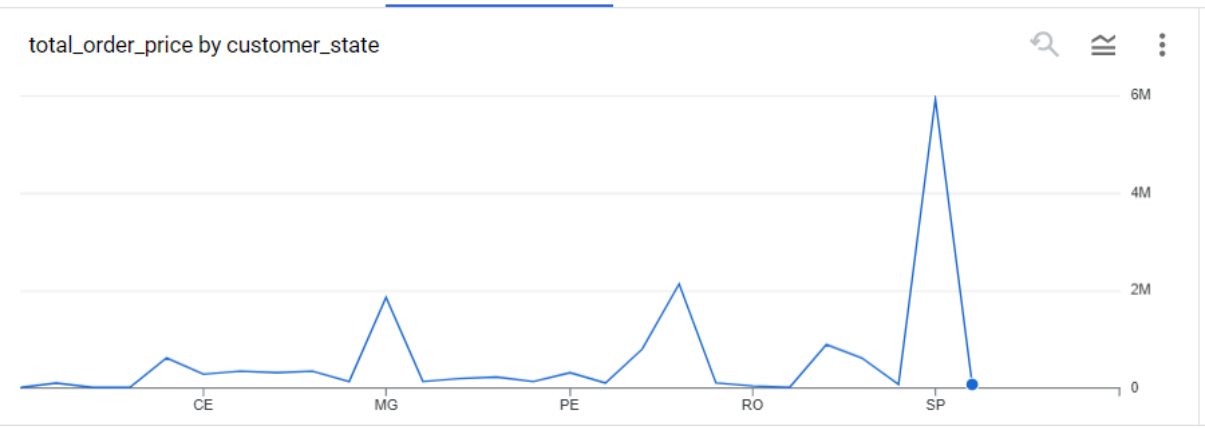
*The above order price is calculated only on the order price value without adding the freight value. To include the freight value in the calculation of the average and total order prices, you can modify the query as follows:

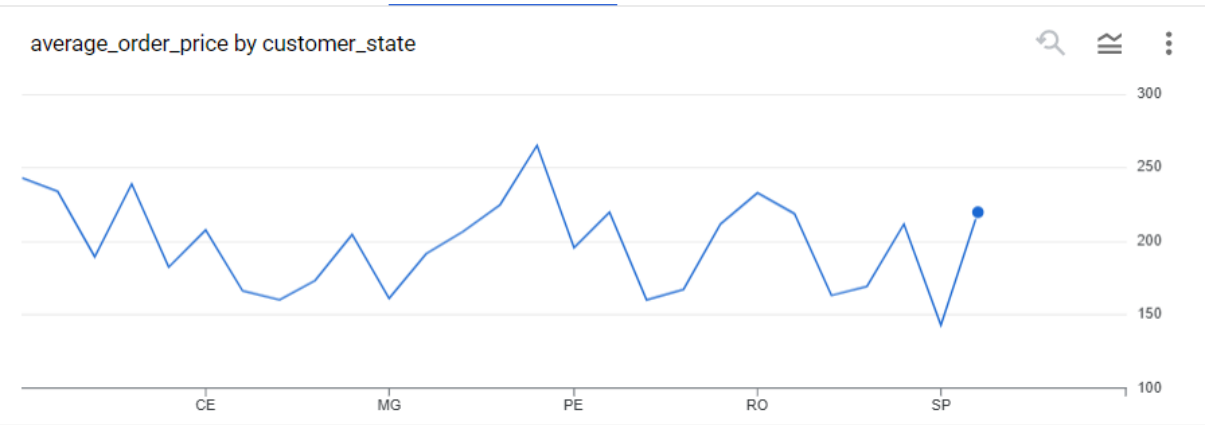
4.2.5. Ans:

```
WITH OrderPrices AS (
  SELECT
    c.customer_state,
    oi.order_id,
    SUM(oi.price + oi.freight_value) AS order_total_price
  FROM
    `Target.customers` c
    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
    c.customer_state, oi.order_id
)

SELECT
  customer_state,
  ROUND(SUM(order_total_price), 2) AS total_order_price,
  ROUND(AVG(order_total_price), 2) AS average_order_price
FROM
  OrderPrices
GROUP BY
  customer_state
ORDER BY
  customer_state;
```

4.2.6. Screenshots:

Row	customer_state	total_order_price	average_order_price
1	AC	19669.7	242.84
2	AL	96229.4	234.13
3	AM	27835.73	189.36
4	AP	16262.8	239.16
5	BA	611506.67	182.1
6	CE	275606.3	207.69
7	DF	353229.44	166.23
8	ES	324801.91	160.4
9	GO	347706.93	173.25
10	MA	151171.99	204.29





Insights and **recommendation** are quite same as above(without including freight prices).

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

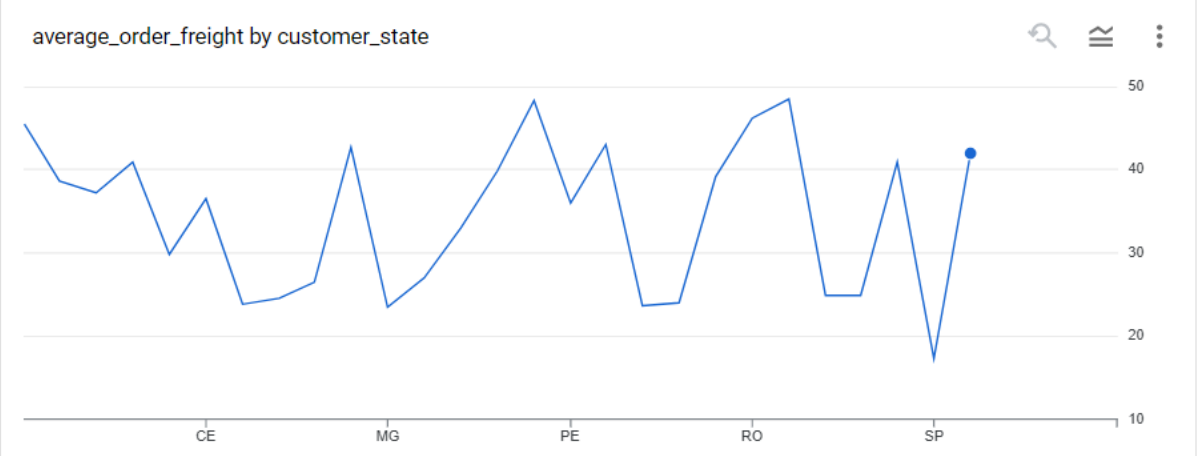
4.3.1. Ans:

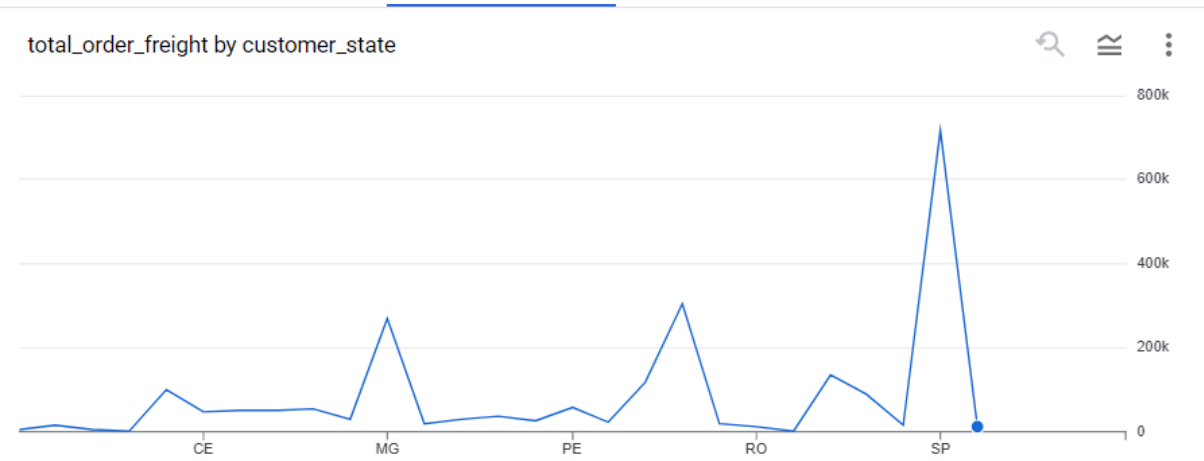
```
WITH OrderFreight AS (
  SELECT
```

```sql
    c.customer_state,
    oi.order_id,
    SUM(oi.freight_value) AS order_total_freight
  FROM
    `Target.customers` c
    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
    c.customer_state, oi.order_id
)

SELECT
  customer_state,
  ROUND(SUM(order_total_freight), 2) AS total_order_freight,
  ROUND(AVG(order_total_freight), 2) AS average_order_freight
FROM
  OrderFreight
GROUP BY
  customer_state
ORDER BY
  customer_state;
```

4.3.2. Screenshots:

Row	customer_state	total_order_freight	average_order_freigh
1	AC	3686.75	45.52
2	AL	15914.59	38.72
3	AM	5478.89	37.27
4	AP	2788.5	41.01
5	BA	100156.68	29.83
6	CE	48351.59	36.44
7	DF	50625.5	23.82
8	ES	49764.6	24.58
9	GO	53114.98	26.46
10	MA	31523.77	42.6





4.3.3. Insights:

4.3.3.1. Average Order Freight Distribution:

- The average order freight values vary across states, ranging from approximately 17.37 to 48.59 units of currency.
- São Paulo (SP) has the lowest average order freight, while Paraíba (PB) has the highest.

4.3.3.2. Total Order Freight Contribution*:*

- São Paulo (SP) contributes significantly to the total order freight, with the highest total order freight among all states.
- Other major contributors include Minas Gerais (MG), Rio de Janeiro (RJ), and Bahia (BA).

4.3.4. Recommendations:

4.3.4.1. Optimize Logistics in High-Freight States:

- In states with higher average order freight values (e.g., PB, PI, AC), consider optimizing logistics and shipping strategies to potentially reduce freight costs.

4.3.4.2. *Evaluate Shipping Partnerships:*

- Explore partnerships with shipping providers to negotiate better freight rates, especially in states where freight costs are relatively high.

4.3.4.3. *Customer Communication:*

- Clearly communicate shipping costs to customers, especially in states with higher average order freight, to manage customer expectations and enhance transparency.

4.3.4.4. *Offer Shipping Discounts:*

- Consider offering shipping discounts or promotions in states where average order freight is relatively high. This can incentivize customers to place more orders.

4.3.4.5. *Efficiency in High-Order Volume States:*

- In states contributing significantly to the total order freight (e.g., SP, MG), focus on optimizing logistics processes to handle high order volumes efficiently.

4.3.4.6. *Continuous Monitoring:*

- Regularly monitor shipping and logistics performance in each state and adjust strategies based on changing market conditions, carrier rates, and customer preferences.

4.3.4.7. *Supply Chain Optimization:*

- Evaluate the supply chain and distribution network to identify opportunities for efficiency improvements, potentially reducing freight costs.

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.

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

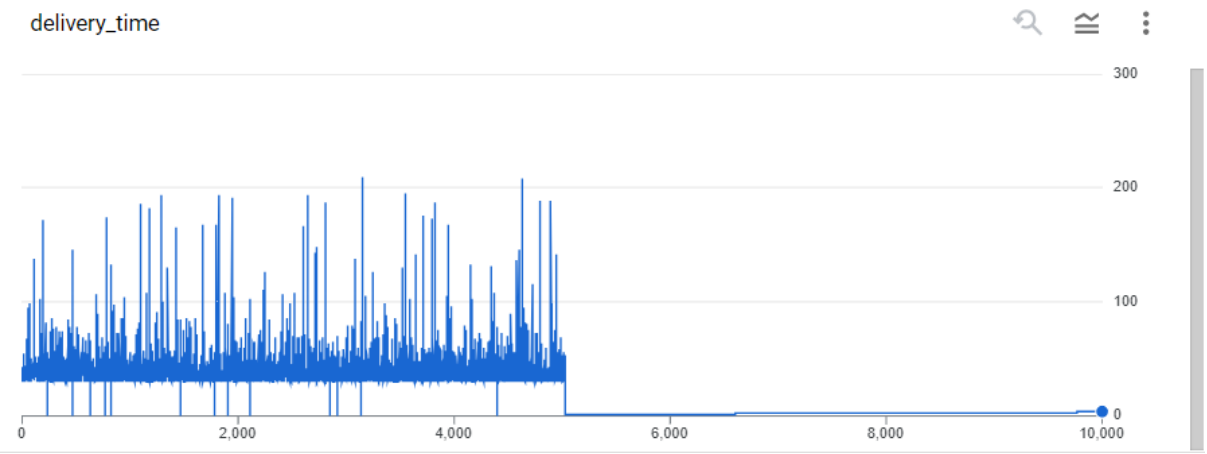
5.1.1. Ans:

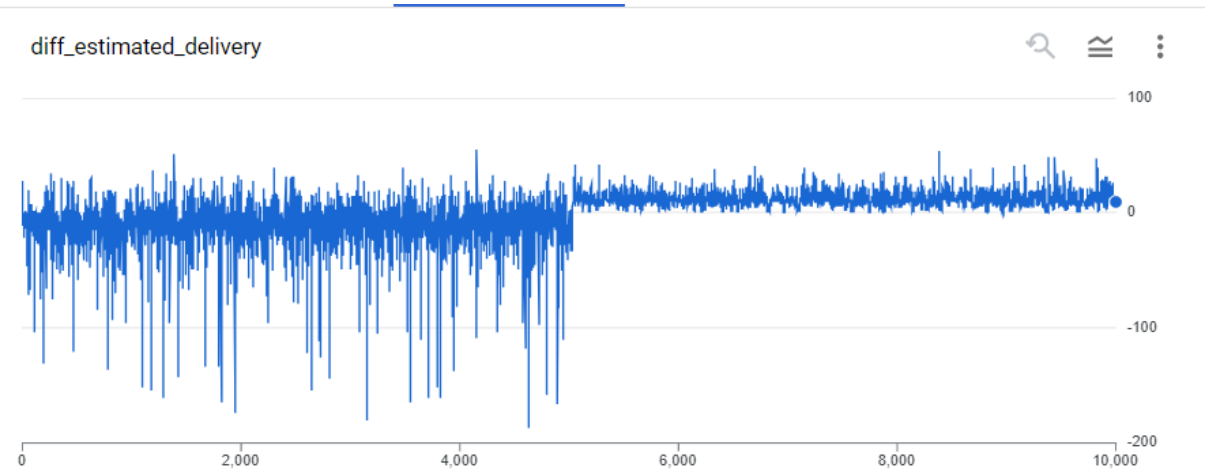
```
SELECT
  order_id,
```

```
  DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) AS delivery_time,
  DATE_DIFF(order_estimated_delivery_date, order_delivered_customer_date, DAY) AS
diff_estimated_delivery
FROM
  `Target.orders`
WHERE
  order_delivered_customer_date IS NOT NULL
  AND order_estimated_delivery_date IS NOT NULL;
```

5.1.2. Screenshots:

Row	order_id	delivery_time	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





5.1.3. Insights:

5.1.3.1. Initial Delivery Time Trend:

- In the initial phase (first 5000 orders), delivery times were relatively high, taking approximately 30+ days. This could indicate challenges in the logistics or operational processes.

5.1.3.2. Sudden Improvement in Delivery Time:

- There is a significant and sudden improvement in delivery time, possibly indicating an infrastructure boost, optimization in logistics, or the implementation of more efficient processes.

5.1.3.3. Shift to 1-Day Delivery:

- As order volume increased, there was a transition from longer delivery times to 1-day delivery, suggesting an enhancement in operational efficiency and fulfillment processes.

5.1.3.4. Consistency in 1-2 Day Delivery:

- The delivery time consistently stayed within the range of 1 to 2 days, showcasing a sustained improvement in the delivery process.

5.1.3.5. Change in Estimated vs. Actual Delivery Difference:

- Initially, the difference between estimated and actual delivery time varied significantly, ranging from +5 to -156 days. This wide range may indicate challenges in accurately estimating delivery times.

5.1.3.6. Improved Estimated vs. Actual Delivery Difference:

- Following the sudden change, there is a notable improvement in the difference between estimated and actual delivery times, with the range now reduced to 0 to +2 days. This indicates a more accurate estimation and better alignment with actual delivery times.

5.1.4. Recommendations:

5.1.4.1. Continue Optimization Efforts:

- Maintain and continue the optimization efforts that led to the sudden improvement in delivery times. This could include investing in technology, improving logistics, or streamlining fulfillment processes.

5.1.4.2. Regular Performance Monitoring:

- Implement regular monitoring of delivery performance to identify any deviations from the 1-2 day delivery range. This proactive approach can help address potential issues promptly.

5.1.4.3. Enhance Estimated Delivery Accuracy:

- Focus on improving the accuracy of estimated delivery times to build trust with customers. This may involve leveraging data analytics and machine learning algorithms to enhance forecasting.

5.1.4.4. Customer Communication:

- Communicate the improved delivery times to customers to enhance their satisfaction and confidence in the brand. Transparency in delivery expectations is crucial for a positive customer experience.

5.1.4.5. Evaluate Infrastructure Investments:

- Assess the impact of any infrastructure investments made during the observed period. Determine the effectiveness of these investments and identify areas for further enhancement.

5.1.4.6. Adapt to Growing Order Volume:

- Continuously adapt operational processes to accommodate the growing order volume. This may involve scaling up resources, optimizing warehouse operations, and collaborating with reliable logistics partners.

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

5.2.1. Ans:

```
WITH StateFreight AS (
  SELECT
    c.customer_state,
    oi.freight_value
```

```
  FROM
    `Target.customers` c
    JOIN `Target.orders` o ON c.customer_id = o.customer_id
    JOIN `Target.order_items` oi ON o.order_id = oi.order_id
)

-- the top 5 states with the highest average freight value
(SELECT
  customer_state,
  ROUND(AVG(freight_value), 2) AS average_freight_value
FROM
  StateFreight
GROUP BY
  customer_state
ORDER BY
  average_freight_value DESC
LIMIT 5)

UNION DISTINCT

-- the top 5 states with the Lowest average freight value
(SELECT
  customer_state,
  ROUND(AVG(freight_value), 2) AS average_freight_value
FROM
  StateFreight
GROUP BY
  customer_state
ORDER BY
  average_freight_value ASC
LIMIT 5)
```

5.2.2. Screenshots:

Row	customer_state	average_freight_valu
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

5.2.3. Insights:

- The states with the highest average freight values are located across different regions, indicating regional variations in freight costs.
- Roraima (RR) and Paraíba (PB) have the highest average freight values, suggesting potential logistical challenges or longer distances.
- São Paulo (SP), being an economic hub, has the lowest average freight value, which may be attributed to better infrastructure and distribution networks.
- States in the Northern and Northeastern regions tend to have higher average freight values compared to the Southern and Southeastern regions.

5.2.4. Recommendations:

5.2.4.1. Optimize Logistics in High-Freight States:

- Explore opportunities to optimize logistics and distribution networks in states with higher average freight values to potentially reduce costs.

5.2.4.2. Evaluate Shipping Partnerships:

- Collaborate with shipping partners to negotiate better freight rates, especially in states where freight costs are relatively high.

5.2.4.3. Adjust Pricing Strategies:

- Consider adjusting product pricing strategies to account for variations in freight costs across different states.

5.2.4.4. Enhance Customer Communication:

- Clearly communicate shipping costs to customers, especially in states with higher average freight values, to manage customer expectations and enhance transparency.

5.2.4.5. Monitor and Adapt:

- Regularly monitor freight costs and adapt strategies based on market conditions, carrier rates, and any changes in logistical dynamics.

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

5.3.1. Ans:

```
WITH StateDeliveryTime AS (
  SELECT
    c.customer_state,
    DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY) AS
delivery_time
  FROM
    `Target.customers` c
    JOIN `Target.orders` o ON c.customer_id = o.customer_id
  WHERE
    o.order_delivered_customer_date IS NOT NULL
```

```
)

(SELECT
  customer_state,
  ROUND(AVG(delivery_time), 2) AS average_delivery_time
FROM
  StateDeliveryTime
GROUP BY
  customer_state
ORDER BY
  average_delivery_time DESC
LIMIT 5)

UNION DISTINCT

(SELECT
  customer_state,
  ROUND(AVG(delivery_time), 2) AS average_delivery_time
FROM
  StateDeliveryTime
GROUP BY
  customer_state
ORDER BY
  average_delivery_time ASC
LIMIT 5)
```

5.3.2. Screenshots:

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

5.3.3. Insights:

- São Paulo (SP), despite having the highest average delivery time among the top 5, still maintains a relatively quick delivery process.
- States in the Southern and Southeastern regions, such as Paraná (PR), Minas Gerais (MG), and Santa Catarina (SC), have faster average delivery times.

- Northern states like Roraima (RR) and Amapá (AP) experience longer average delivery times, possibly due to logistical challenges or geographical distances.
- The overall average delivery times are within a reasonable range, indicating a generally efficient delivery process.

5.3.4. Recommendations:

5.3.4.1. Optimize Logistics in Northern States:

- Explore opportunities to optimize logistics and distribution networks in states with longer average delivery times, focusing on Northern regions.

5.3.4.2. Enhance Communication:

- Clearly communicate expected delivery times to customers, especially in states with longer average delivery times, to manage expectations and improve customer satisfaction.

5.3.4.3. Evaluate Shipping Partnerships:

- Collaborate with shipping partners to identify opportunities for faster and more reliable deliveries, especially in regions with longer average delivery times.

5.3.4.4. Monitor and Adapt:

- Regularly monitor delivery times, identify trends, and adapt strategies based on customer feedback and market dynamics.

5.4. Find out the top 5 states where the order delivery is 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.

5.4.1. Ans:

```
WITH StateDeliverySpeed AS (
  SELECT
    c.customer_state,
    DATE_DIFF(o.order_delivered_customer_date, o.order_estimated_delivery_date, DAY) AS
delivery_speed
  FROM
    `Target.customers` c
    JOIN `Target.orders` o ON c.customer_id = o.customer_id
  WHERE
    o.order_delivered_customer_date IS NOT NULL
)

SELECT
  customer_state,
  ROUND(AVG(delivery_speed), 2) AS average_delivery_speed
FROM
  StateDeliverySpeed
```

```
GROUP BY
  customer_state
ORDER BY
  average_delivery_speed ASC
LIMIT 5;
```

5.4.2. Screenshots:

Row	customer_state	average_delivery_spe
1	AC	-19.76
2	RO	-19.13
3	AP	-18.73
4	AM	-18.61
5	RR	-16.41

5.4.3. Insights:

- These states consistently show negative values for average delivery speed, indicating that, on average, orders are delivered before the estimated date.
- The negative values suggest a trend of faster order fulfillment, possibly due to efficient logistics, shorter delivery distances, or streamlined operations.

5.4.4. Recommendations:

5.4.4.1. Analyze Operational Efficiency:

- Investigate the operational processes in these states to understand what contributes to the faster order delivery. Identify and implement best practices across other regions.

5.4.4.2. Communicate Realistic Estimated Delivery Dates:

- Ensure that estimated delivery dates provided to customers align with the actual capabilities of the logistics network. Consider adjusting estimated dates based on the performance observed in these states.

5.4.4.3. Share Success Stories:

- Use the positive delivery speed performance in marketing and customer communication to build trust and attract more customers.

5.4.4.4. Explore Optimization Opportunities:

- Continue exploring opportunities for optimization, even in states with positive delivery speed, to maintain a consistent and efficient delivery experience.

6. Analysis based on the payments:

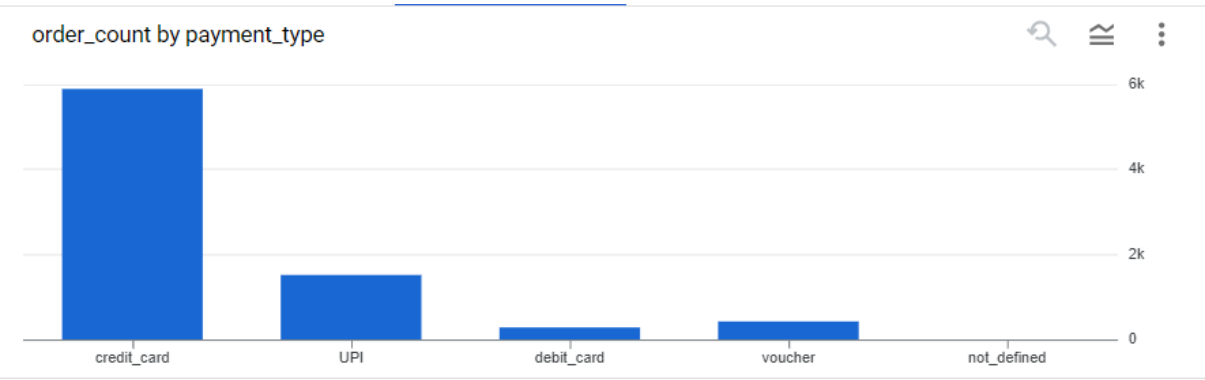
6.1. Find the month on month no. of orders placed using different payment types.

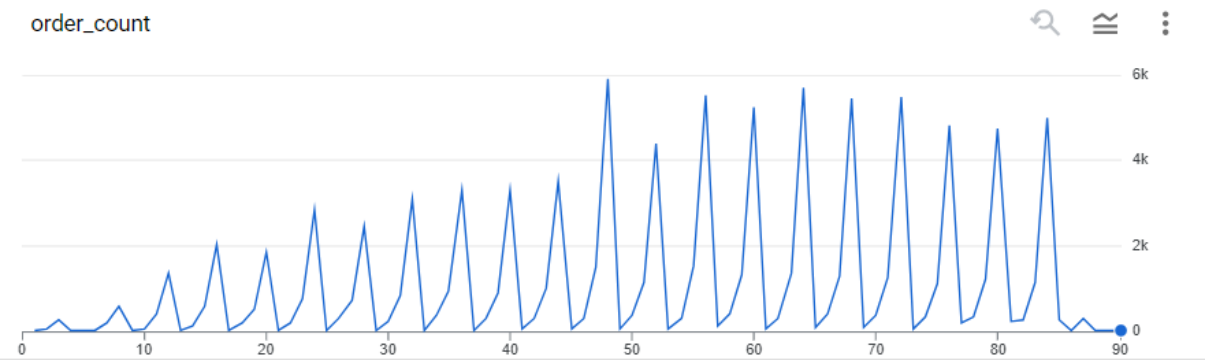
6.1.1. Ans:

```
WITH MonthlyOrderCounts AS (
  SELECT
    EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
    p.payment_type,
    COUNT(*) AS order_count
  FROM
    `Target.orders` o
    JOIN `Target.payments` p ON o.order_id = p.order_id
  GROUP BY
    order_year, order_month, payment_type
  ORDER BY
    order_year, order_month
)

SELECT
  order_year,
  order_month,
  payment_type,
  COALESCE(order_count, 0) AS order_count
FROM
  MonthlyOrderCounts
ORDER BY
  order_year, order_month, payment_type;
```

6.1.2. Screenshots:

Row	order_year	order_month	payment_type	order_count
1	2016	9	credit_card	3
2	2016	10	UPI	63
3	2016	10	credit_card	254
4	2016	10	debit_card	2
5	2016	10	voucher	23
6	2016	12	credit_card	1
7	2017	1	UPI	197
8	2017	1	credit_card	583
9	2017	1	debit_card	9
10	2017	1	voucher	61





6.1.3. Insights:

6.1.3.1. Payment Type Distribution:

- The most commonly used payment types are credit cards, UPI (Unified Payments Interface), debit cards, and vouchers.
- Credit cards and UPI are the predominant payment methods, with consistent usage throughout the observed period.

6.1.3.2. Monthly Trends:

- There is a general upward trend in the number of orders placed using various payment types over the months.
- Months with higher sales, such as November and December, show increased usage of all payment types.

6.1.3.3. Credit Card Dominance:

- Credit cards consistently contribute to the majority of payments across all months, indicating their popularity among customers.

6.1.3.4. Seasonal Impact:

- Voucher usage shows some seasonality, with peaks during certain months, possibly influenced by promotional campaigns or festive seasons.

6.1.3.5. Decline in Voucher Usage:

- The usage of vouchers declines in the later months of 2018, indicating a potential shift in customer behavior or changes in promotional strategies.

6.1.4. Recommendations:

6.1.4.1. Promote Credit Card Offers:

- Given the consistent dominance of credit card payments, promotions or discounts related to credit card usage could attract more customers.

6.1.4.2. Monitor and Optimize Voucher Campaigns:

- Keep a close eye on voucher usage trends and optimize promotional campaigns accordingly. Identify factors influencing the decline in voucher usage and adjust strategies.

6.1.4.3. Enhance UPI Integration:

- UPI shows consistent growth and usage. Consider enhancing UPI integration and exploring partnerships to capitalize on its popularity.

6.1.4.4. Seasonal Promotions:

- Leverage seasonal trends by introducing targeted promotions or discounts during peak months to maximize sales.

6.1.4.5. Customer Education:

- Educate customers about the benefits of different payment methods, encouraging them to explore and use a variety of options.

6.1.4.6. Adapt to Changes:

- Stay adaptable to changes in customer behavior and preferences. Regularly analyze payment data to identify emerging trends and make data-driven decisions.

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

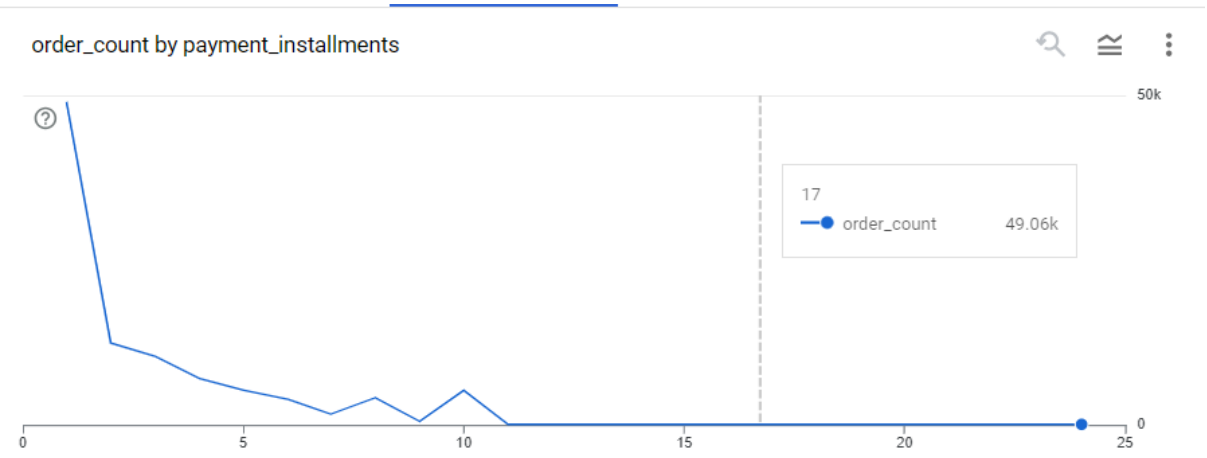
6.2.1. Ans:

```
SELECT
  payment_installments,
  COUNT(DISTINCT order_id) AS order_count
FROM
```

```
  `Target.payments`
WHERE
  payment_installments > 0
GROUP BY
  payment_installments
ORDER BY
  payment_installments;
```

6.2.2. Screenshots:

Row	payment_installment	order_count
1	1	49060
2	2	12389
3	3	10443
4	4	7088
5	5	5234
6	6	3916
7	7	1623
8	8	4253
9	9	644
10	10	5315



6.2.3. Insights:

- Most orders have 1 or 2 payment installments, indicating that a significant number of customers prefer single or two-part payments.
- There is a gradual decline in the number of orders as the number of payment installments increases.
- Installments beyond 10 show a relatively lower frequency, suggesting that extended payment plans are less common.

6.2.4. Recommendations:

6.2.4.1. Promote Flexible Payment Options:

- Highlight and promote payment options with 1 or 2 installments, as they are preferred by the majority of customers.
- Consider offering special promotions or discounts for customers choosing shorter installment plans.
- Communicate Benefits of Larger Installments:

- Educate customers about the benefits or perks of larger installments, such as reduced financial burden or potential discounts for bulk payments.

6.2.4.2. Review High Installment Numbers:

- Investigate the reason behind a small number of orders with high installment counts (e.g., 11 or more). Understand customer preferences and potential barriers to choosing longer installment plans.

6.2.4.3. Enhance Customer Education:

- Provide clear and transparent information about installment options during the checkout process to help customers make informed decisions.

6.2.4.4. Explore Customized Payment Plans:

- Consider offering customizable payment plans based on customer preferences and financial situations.