```
-- 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.
SELECT *
FROM `case_data.customers`;
```

Row	ustomer_id	customer_unique_id	customer_zip_code_	customer_city	customer_state
1	735e7e4298a2ebbb46649346...	fcb003b1bdc0df64b4d065d9b...	59650	acu	RN
2	03b3d86e3990db01619a4ebe...	46824822b15da44e983b021d...	59650	acu	RN
3	3c97666e962d4fea7fd6a83e...	b6108acc674ae5c99e29adc10...	59650	acu	RN
4	7c2f46cf580f4874c9a5751c2...	402cce5c0509000eed9e77fec...	63430	ico	CE
5	13ef4cfffb8ad4767c199c36a...	6ba00666ab7eada5ceec279b2...	63430	ico	CE
6	00841b86e1fbe9493b52324...	796a0b1a21f597704057184a1...	63430	ico	CE
7	325415ccc7e622c66dec4bc...	05d1d2d9f0161c5f397ce7fc77...	63430	ico	CE
8	1f3a7b250e3be964f01bf22bc...	c34585a0276ecc5e4fb03de75...	63430	ico	CE
9	04202b8ef01f4719a4691e79...	01a4fe5fc00bbdb0b0a4af5a53...	63430	ico	CE

We ensure appropriate analysis and interpretation of the dataset by comprehending the data kinds of each table.

```
-- 1.2. Get the time range between which the orders were placed.
SELECT
MIN(order_purchase_timestamp), MAX(order_purchase_timestamp)
FROM `case_data.orders`;
```

Row	f0_	f1_
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

```
-- 1.3. Count the Cities & States of customers who ordered during the given period.
SELECT
COUNT(DISTINCT customer_city) AS total_city, COUNT(DISTINCT customer_state) AS
total_state
FROM `case_data.customers` c
INNER JOIN `case_data.orders` o
ON c.customer_id = o.customer_id;
```

Row	total_city	total_state
1	4119	27

```
-- 2. In-depth Exploration:
-- 2.1. Is there a growing trend in the no. of orders placed over the past years?
SELECT
  EXTRACT(YEAR FROM o.order_purchase_timestamp) AS year,
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
  COUNT(DISTINCT o.order_id) AS order_count
FROM
  `case_data.orders` o
JOIN
  `case_data.customers` c
ON
  o.customer_id = c.customer_id
GROUP BY
  year, month
ORDER BY
  year, month;
```

Row	year	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

It is clear that e-commerce is becoming more popular in Brazil. With minor variations, the number of purchases has demonstrated a general rising tendency.

Here I can see there is some fluctuation, so we need to find the reason behind it and fix them as soon as possible so that it will lead to the overall growth of the business.

```
--2.2. Can we see some kind of monthly seasonality in terms of the no. of orders
being placed?
  SELECT
  EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
  COUNT(DISTINCT order_id) AS order_count
FROM
  `case_data.orders`
GROUP BY
  month
ORDER BY
  month;
```

Row	month	order_count
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843

Making definite judgments about seasonality trends is difficult. However, we can see some seasonality in the e-commerce orders through the analysis and visualization in big query There are peaks and valleys in the number of orders from March to August. Notably, there is a rise in orders in February and March, which falls around Brazil's Carnival season.

```sql
 -- 2.3. During what time of the day, do the Brazilian customers mostly place their
orders? (Dawn, Morning, Afternoon or Night)
  SELECT
  CASE
    WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 0 AND 5 THEN 'Dawn'
    WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 6 AND 11 THEN
'Morning'
    WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 12 AND 17 THEN
'Afternoon'
    WHEN EXTRACT(HOUR FROM o.order_purchase_timestamp) BETWEEN 18 AND 23 THEN
'Night'
  END AS hour,
  COUNT(o.order_id) AS order_count
FROM
  `case_data.orders` o
JOIN
  `case_data.customers` c
ON o.customer_id = c.customer_id
GROUP BY
  hour
ORDER BY
  order_count DESC;
```

Row	hour	order_count
1	Afternoon	38361
2	Night	34100
3	Morning	22240
4	Dawn	4740

The data revealed that Brazilian clients often place the majority of their orders during the day, particularly in the late afternoon and evening. This shows that users prefer to purchase online during their free time or after finishing their everyday tasks. It's crucial to understand that the recorded timestamps are presumed to reflect the right time zone at the moment of purchase.

E-commerce companies are able to run their operations more efficiently by comprehending client purchasing behaviour. Companies may more efficiently distribute resources, such customer service agents and merchandise, to fulfil consumer requests and deliver a smooth shopping experience by detecting peak purchase hours.

```sql
  -- 3. Evolution of E-commerce orders in the Brazil region.
  -- 3.1. Get the month on month no. of orders placed in each state.
SELECT
  c.customer_state,
  EXTRACT(month FROM o.order_purchase_timestamp) AS month,
  COUNT(o.order_purchase_timestamp) AS order_count
FROM
  `case_data.orders` o
JOIN
  `case_data.customers` c
ON
  o.customer_id = c.customer_id
GROUP BY
  c.customer_state, month
  ORDER BY
  c.customer_state, month;
```

Row	customer_state	month	order_count
3	AC	3	4
4	AC	4	9
5	AC	5	10
6	AC	6	7
7	AC	7	9
8	AC	8	7
9	AC	9	5
10	AC	10	6

It displays the month-over-month order numbers for each state in Brazil, giving important details on regional client shopping patterns.

```
 -- 3.2. How are the customers distributed across all the states?
  SELECT
  c.customer_state,
  COUNT(c.customer_id) AS no_of_customers
FROM
  `case_data.customers` c
GROUP BY
  c.customer_state
ORDER BY
  no_of_customers DESC;
```

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

The data shows that the state of Sao Paulo (SP), which is the most populated state in Brazil, has the greatest number of consumers. This result is consistent with the earlier research and shows a positive association between a state's population and order count.

It is essential for companies like Target to comprehend the growth of e-commerce orders and the distribution of customers throughout Brazilian states in order to customise their marketing strategies, optimise logistics, and improve customer experiences. By utilising this SQL-driven research, e-commerce businesses are better able to target certain geographic areas, spend resources wisely, and provide tailored experiences that meet the distinct tastes and needs of clients in various states.

```sql
  -- 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).
SELECT
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
  (
    (
      SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018 AND
      EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 THEN
      p.payment_value END)
      -
      SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 AND
      EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 THEN
      p.payment_value END)
    )
    /
    SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 AND
    EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8 THEN
    p.payment_value END)
  )*100 AS percent_increase
FROM
  `case_data.orders` o
JOIN
  `case_data.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
GROUP BY 1
ORDER BY 1;
```

Row	month	percent_increase
1	1	705.1266954171...
2	2	239.9918145445...
3	3	157.7786066709...
4	4	177.8407701149...
5	5	94.62734375677...
6	6	100.2596912456...
7	7	80.04245463390...
8	8	51.60600520477...

Including only the months of January through August, the overall percentage rise in order costs from 2017 to 2018 is 138.53%. The largest percentage rise can be seen when looking at the growth month-by-month; it is followed by February and April.

```sql
-- 4.2. Calculate the Total & Average value of order price for each state.
-- 4.3. Calculate the Total & Average value of order freight for each state.
SELECT
  c.customer_state,
  ROUND(AVG(i.price), 2) AS mean_price,
  ROUND(SUM(i.price), 2) AS total_price,
  ROUND(AVG(i.freight_value), 2) AS mean_freight_value,
```

```
  ROUND(SUM(i.freight_value), 2) AS total_freight_value
FROM
  `case_data.orders` o
JOIN
  `case_data.order_items` i ON o.order_id = i.order_id
JOIN
  `case_data.customers` c ON o.customer_id = c.customer_id
GROUP BY
  c.customer_state;
```

Row	customer_state	mean_price	total_price	mean_freight_value	total_freight_value
1	MT	148.3	156453.53	28.17	29715.43
2	MA	145.2	119648.22	38.26	31523.77
3	AL	180.89	80314.81	35.84	15914.59
4	SP	109.65	5202955.05	15.15	718723.07
5	MG	120.75	1585308.03	20.63	270853.46
6	PE	145.51	262788.03	32.92	59449.66
7	RJ	125.12	1824092.67	20.96	305589.31
8	DF	125.77	302603.94	21.04	50625.5
9	RS	120.34	750304.02	21.74	135522.74
10	SE	153.04	58920.85	36.65	14111.47

Interesting conclusions are drawn from the analysis. Despite having the greatest total price value and total freight value among all states, So Paulo (SP) has the lowest average price value and average freight value. The state of Paraba (PB), on the other side, has the greatest average freight value and average pricing value.

An extensive examination of cost patterns, price movements, and freight values is necessary to comprehend the effects on the economy. Businesses may obtain important insights into the economic environment, spot possible growth opportunities, and make data-driven decisions to optimise pricing strategies, improve logistics, and increase overall economic effect by applying SQL queries and looking at state-wise trends.

```
--6. Analysis based on the payments:

--6.1. Find the month on month no. of orders placed using different payment types.
SELECT
  p.payment_type,
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
  COUNT(DISTINCT o.order_id) AS order_count
FROM
  `case_data.orders` o
JOIN
  `case_data.payments` p
ON
  o.order_id = p.order_id
GROUP BY
  1, 2
ORDER BY
  1, 2;
```

Row	payment_type	month	order_count
3	UPI	3	1942
4	UPI	4	1783
5	UPI	5	2035
6	UPI	6	1807
7	UPI	7	2074
8	UPI	8	2077
9	UPI	9	903
10	UPI	10	1056

The study reveals two uptrends: one from September to November and one from January to August. The most common payment method is credit card transactions, followed by UPI. Transactions made with a debit card are the least recommended. Notably, compared to other payment methods, credit card transactions are growing quickly. This trend may be attributed to advantages like "buy now, pay later" possibilities or cashback rewards offered by credit cards.

```
--6.2. Find the no. of orders placed on the basis of the payment installments that have been paid.
SELECT
  p.payment_installments,
  COUNT(o.order_id) AS order_count
FROM
  `case_data.orders` o
JOIN
  `case_data.payments` p
ON
  o.order_id = p.order_id
WHERE
  o.order_status != 'canceled'
GROUP BY
  1
ORDER BY
  2 DESC;
```

Row	payment_installment	order_count
1	1	52184
2	2	12353
3	3	10392
4	4	7056
5	10	5292
6	5	5209
7	8	4239
8	6	3898
9	7	1620
10	9	638

According to the data, the majority of orders only have one payment installment (at the most). The most installments are 24, which are connected to 18 orders.

For businesses to optimize their payment processes and satisfy client preferences, they must have a thorough understanding of payment forms and installment preferences. Companies may improve payment alternatives, optimize operations, and improve the entire consumer experience by using SQL queries and analyzing payment statistics.

Insights & Recommendations based on this analysis:

According to the data, SP has a lot more orders than the following five states put together. This suggests that there is room for advancement in the other states. By concentrating on these states, you may boost sales and grow your clientele.

Sales fluctuate seasonally, with holiday seasons seeing the biggest increases. To take advantage of these peak times and improve customer satisfaction, businesses can organize their marketing and sales activities accordingly. This will lead to an increase in overall revenues.

States with high order counts include SP and RJ. It is advised to concentrate on client retention techniques, such as tailored marketing campaigns, loyalty programs, and first-rate customer care, in order to increase sales and create brand loyalty.

Analyzing consumer demographics can offer insightful information for developing goods and marketing plans that are specifically targeted to target markets. Sales and customer satisfaction may rise as a result of this personalization.

The data shows that orders decreased in September and October. By providing discounts or other incentives during off-peak times, businesses may encourage customers to make purchases and increase sales.

Economic circumstances are not included in the statistics, but analyzing how they affect sales can assist identify areas for investment and development, guaranteeing resilience amid economic swings.

Recommendations:

Enhance shipping and logistics procedures to shorten delivery times and boost customer satisfaction. This entails streamlining shipping processes, enhancing warehouse operations, and collaborating with reputable courier firms.

Implement customer retention methods to promote loyalty and repeat business. Personalized offers, referral bonuses, and loyalty programs may all help you do this.

To maintain market competitiveness while maximizing revenue and profitability, evaluate pricing and freight charges. As necessary, take into account raising pricing or modifying freight charges.

Invest in infrastructure and technology to improve the online shopping experience. This includes putting chatbots to use for customer service, enhancing website functionality, and making tailored product suggestions depending on user behavior.

Work together with vendors to increase product selection and quality while taking into account the varied demands and preferences of customers.

Utilize social media channels and influencers to market items and raise brand recognition because these factors strongly affect Brazilian consumers' purchase decisions.

By providing chat support services and assuring quick, efficient replies to consumer inquiries, you may improve the customer service experience.

In order to be competitive in the market, keep an eye on rival activities and make necessary adjustments to the company plan, such as matching or improving prices, extending the range of available products, or enhancing customer service.

Businesses may optimize their operations, improve customer happiness, and promote overall sales development in the Brazilian e-commerce industry by putting these practical insights and suggestions into practice.