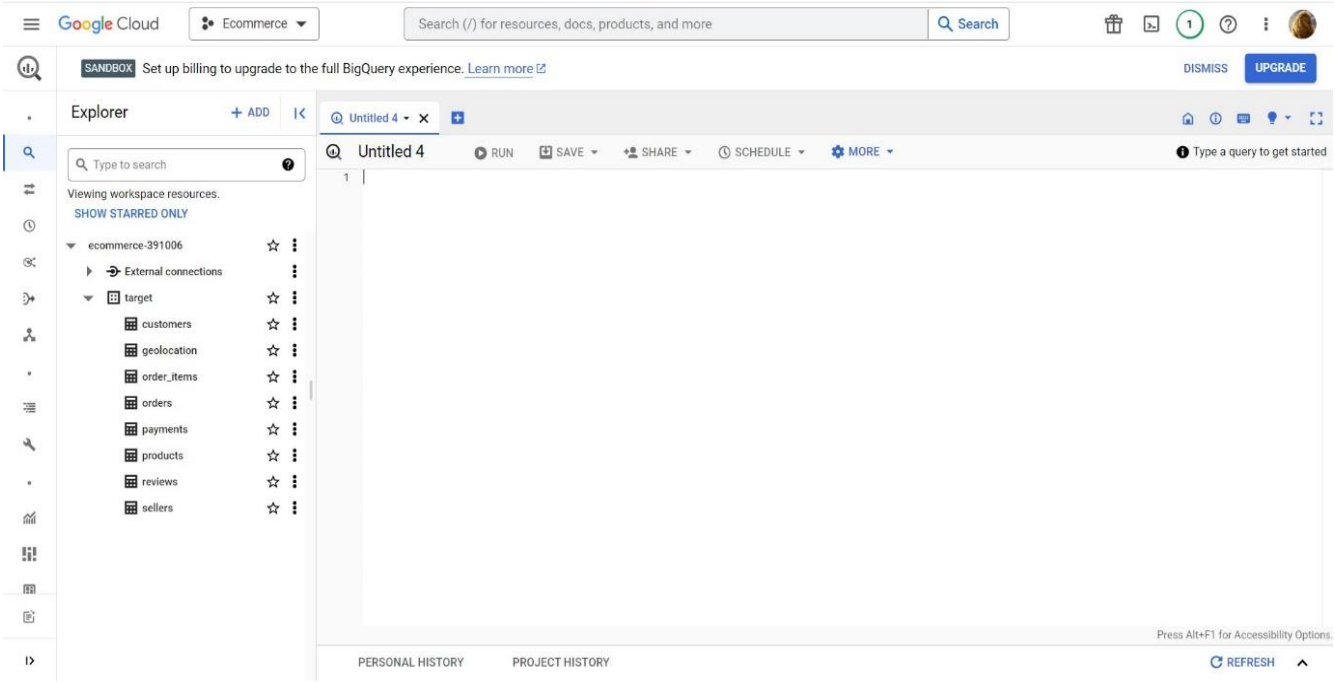
TARGET
BUSINESS CASE STUDY REPORT

LOGESH G

DSML NOV 22 MORNING BATCH

Mail Id: logeshwaran38cgm@gmail.com

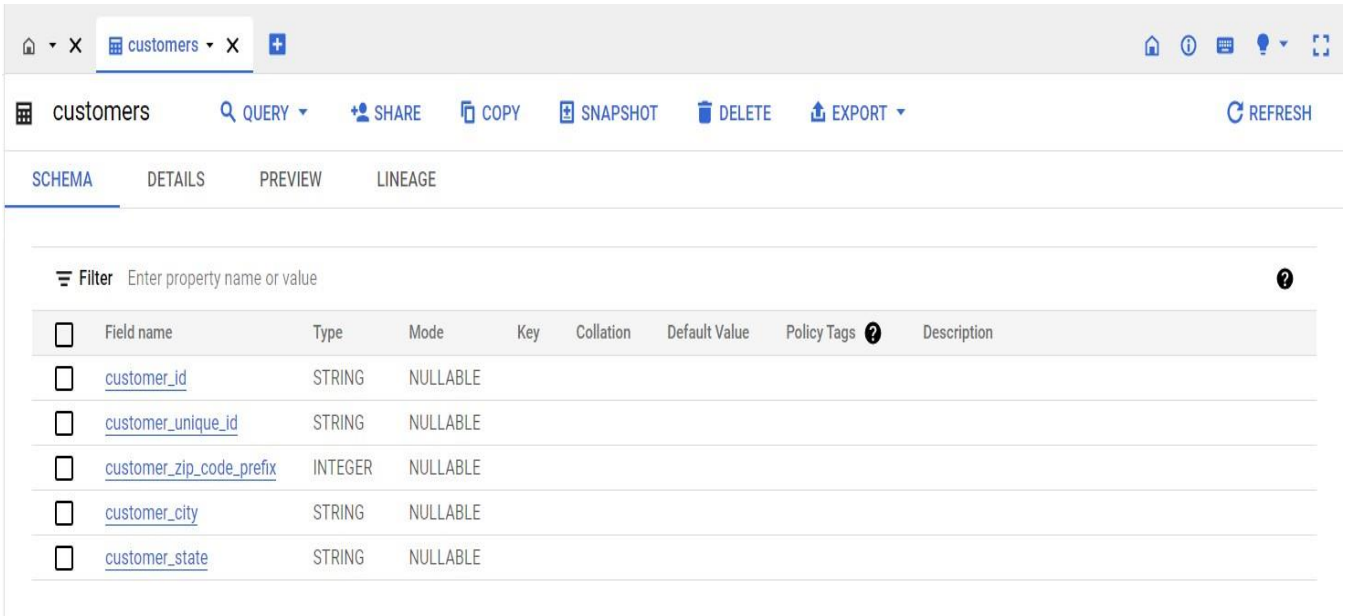
Uploading Dataset and Creating Columns



BUSINESS CASE ANALYSIS

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

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



Insights:

Data Types: It's important to ensure that the data types assigned to each column accurately represent the nature of the data. Based on the given column names, the data types assigned to the columns seem appropriate. However, it's always recommended to cross-verify the actual data in each column to confirm that the data types align with the values.

we have a dataset named "customers" with the following columns:

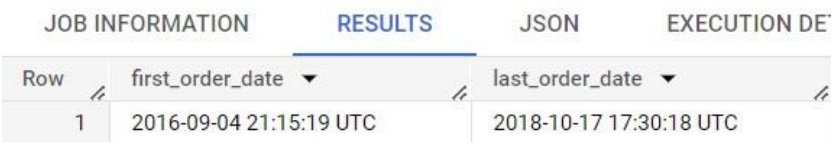
- customer_id:
  The data type of this column is STRING.
- customer_unique_id:
  The data type of this column is STRING.
- customer_zip_code_prefix:
  The data type of this column is INTEGER.
- customer_city:
  The data type of this column is STRING.
- customer_state:
  The data type of this column is STRING.

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

Query:

```
SELECT MIN(order_purchase_timestamp) AS first_order_date,
MAX(order_purchase_timestamp) AS last_order_date
FROM target.orders;
```

Result:



Insights and Recommendations:

Duration of Operations:

- The time range between the first and last order dates indicates the duration of operations for the business.
- In this case, the business operated for approximately 2 years and 1 month.
- This insight provides an understanding of the historical context and timeframe for analyzing business performance.

Business Growth:

- Analyzing the time range can provide insights into the business's growth trajectory.
- By examining the order volume and revenue trends over time, we can identify periods of growth, stability, or decline.
- This information can help in evaluating the business's overall performance and identifying factors contributing to its success or challenges.

Seasonality:

- By further exploring the distribution of orders within different time periods (e.g., months, quarters, or years), which can uncover seasonal patterns.
- Understanding seasonality is crucial for planning marketing campaigns, managing inventory, and optimizing operations to meet increased demand during peak seasons and effectively utilize resources during slower periods.

Identify Peak Seasons:

- Utilize the historical order data to identify peak seasons or periods of high demand.
- This knowledge can inform inventory management, staffing, and marketing strategies to ensure adequate resources are allocated during peak times.

Forecasting and Inventory Management:

- By understanding demand patterns and trends, we can ensure the right products are available at the right time, improving customer satisfaction and operational efficiency.

Business Expansion Opportunities:

- Based on the historical order data, identify geographical areas or customer segments that have shown growth potential.

c. Count the number of Cities and States in our dataset.

Query:

```
SELECT COUNT(DISTINCT customer_city) AS Number_of_cities,
COUNT(DISTINCT customer_state) AS Number_of_states
FROM target.customers;
```

Result:

JOB INFORMATION | RESULTS | JSO

Row	Number_of_cities	Number_of_states
1	4119	27

Insights and Recommendations:

Geographic Distribution:

- The dataset contains customers from 4,119 unique cities.
- This indicates a diverse customer base spread across multiple locations.
- Analyzing the distribution of customers across cities can provide insights into regional preferences, target markets, and potential opportunities for localized marketing or expansion strategies.

State-level Analysis:

- With customers from 27 different states, analyzing customer behaviour and preferences at the state level can reveal patterns and trends.
- Understanding variations in customer demographics, buying habits, and preferences among different states can help tailor marketing campaigns, product offerings, and customer service strategies to meet regional demands.

Concentration of Customers:

- Assessing the concentration of customers across cities and states can help identify areas with a high customer density.
- This information can guide decisions related to resource allocation, distribution networks, and the opening of new physical locations or warehouses.

Customer Segmentation:

- Utilize the geographic information to segment customers based on their city or state.
- This segmentation can be valuable for targeted marketing campaigns, personalized communication, and region-specific offers.

- By understanding the unique characteristics of customers in different cities or states, we can better tailor your marketing strategies and improve customer engagement.

Market Expansion Opportunities:

- Analyzing the dataset's coverage across cities and states can highlight regions that are underrepresented or have untapped potential.
- Identifying areas with a limited customer base can inform decisions related to market expansion, strategic partnerships, and targeted acquisition efforts to capture new customers in those regions.

2. In-depth Exploration:

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

```
SELECT count(*) as orders_per_monthandyear,EXTRACT(YEAR
FROM order_purchase_timestamp) as purchase_year,
EXTRACT(MONTH FROM order_purchase_timestamp) as
purchase_month
from `target.orders`
group by purchase_month,purchase_year
order by purchase_month,purchase_year
```

Result:

Row	orders_per_monthan	purchase_year	purchase_month
1	800	2017	1
2	7269	2018	1
3	1780	2017	2
4	6728	2018	2
5	2682	2017	3
6	7211	2018	3
7	2404	2017	4
8	6939	2018	4
9	3700	2017	5
10	6873	2018	5

Insights and Recommendations:

Comparison between 2017 and 2018:

- January:
  The number of orders in January 2018 (7269) is significantly higher than in January 2017 (800).
  This indicates substantial growth between the two years.
- February:
  In February 2018, there was an increase in the number of orders (6728) compared to February 2017 (1780), indicating growth.
- March:
  Similar to February, there is growth in the number of orders in March 2018 (7211) compared to March 2017 (2682).
- April:
  The number of orders in April 2018 (6939) is higher than in April 2017 (2404), demonstrating growth.
- May:
  May 2018 (6873) shows a higher number of orders compared to May 2017 (3700), indicating growth.

Monthly Patterns:

- The analysis of order counts by month within each year suggests potential seasonality or periodic trends.
- In both 2017 and 2018, there is an upward trend in the number of orders from January to February, followed by a further increase in March and April.
- May shows a similar pattern, but the growth is not as pronounced as in March and April.
- Understanding these monthly patterns can help with inventory management, resource allocation, and planning marketing campaigns to capitalize on periods of higher order volumes.

Overall Growth Trend:

- The data suggests an overall growth trend in the number of orders placed over the past years.
- Comparing the order counts between the corresponding months of 2017 and 2018 consistently reveals higher order volumes in 2018.
- This growth trend indicates a potential increase in customer demand or business expansion during the analysed period.

Factors Influencing Growth:

- The growth in order counts may be influenced by various factors such as marketing initiatives, improvements in customer experience, product diversification, or increased customer base.
- It would be beneficial to explore these factors further to identify key drivers and develop strategies to sustain and enhance the growth trend.

Future Analysis:

- Continuously monitoring and analyzing the order counts over time will provide a clearer picture of the growth trend.
- Consider expanding the analysis to include more recent years to identify any sustained growth or potential fluctuations.
- In-depth analysis involving customer segmentation, product performance, and marketing campaign evaluation can provide deeper insights into the underlying factors contributing to the growth trend.

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

Query:

```
SELECT EXTRACT(MONTH FROM order_purchase_timestamp) AS month, COUNT(*) AS Number_of_orders
FROM target.orders
GROUP BY month
ORDER BY month;
```

Result:

JOB INFORMATION | RESULTS | JSO

Row	month	Number_of_orders
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305
10	10	4959
11	11	7544
12	12	5674

Insights:

- Monthly Seasonality:

  The analysis of the number of orders placed per month suggests the presence of monthly seasonality.
- Peak Months:
  - Months with higher order volumes include March, May, July, and August.
  - These months consistently exhibit higher order counts compared to other months.
- Seasonal Patterns:
  - There is a visible trend of increasing order counts from January (8069) to May (10573), followed by a slight dip in June (9412).
  - Then, there is another increase in order counts from July (10318) to August (10843). September (4305) and October (4959) show lower order counts compared to the peak months.
- Winter Period:
  - The winter months, such as January and February, exhibit relatively lower order counts compared to the peak months.
  - This could be attributed to factors like post-holiday lull or reduced consumer spending during that period.

Recommendations:

Seasonal Marketing Campaigns:

- Capitalize on the months with higher order volumes (March, May, July, August) by planning targeted marketing campaigns, promotions, and product launches during these periods.
- Align your marketing efforts to meet increased customer demand during peak months.

Inventory Management:

- Adjust inventory levels and supply chain operations to align with seasonal patterns. Anticipate higher demand during peak months and ensure sufficient stock availability to avoid stockouts and missed sales opportunities.
- Use demand forecasting techniques to optimize inventory management.

Customer Engagement:

- Leverage seasonal trends to enhance customer engagement.
- Tailor communication and promotional strategies to align with the preferences and needs of customers during specific months.

- Personalize offers and incentives based on historical order patterns and customer segmentation.

Pricing Strategies:

- Analyze price elasticity and demand sensitivity during peak and off-peak months.
- Consider dynamic pricing strategies to optimize revenue and customer satisfaction. Implement promotional pricing during off-peak months to stimulate demand and drive sales.

Market Expansion Opportunities:

- Explore potential expansion into new markets or regions that show different seasonal patterns.
- Identify areas where seasonality trends differ from the current customer base, allowing for market diversification and year-round revenue streams.

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

- **0-6 hrs : Dawn**
- **7-12 hrs : Mornings**
- **13-18 hrs : Afternoon**
- **19-23 hrs : Night**

Query:

```
SELECT
  CASE
       WHEN EXTRACT(HOUR FROM order_purchase_timestamp)
BETWEEN 0 AND 6 THEN 'Dawn'
       WHEN EXTRACT(HOUR FROM order_purchase_timestamp)
BETWEEN 7 AND 12 THEN 'Morning'
       WHEN EXTRACT(HOUR FROM order_purchase_timestamp)
BETWEEN 13 AND 18 THEN 'Afternoon'
       WHEN EXTRACT(HOUR FROM order_purchase_timestamp)
BETWEEN 19 AND 23 THEN 'Night'
  END AS time_of_day,
  COUNT(*) AS Number_of_orders
FROM target.orders
GROUP BY time_of_day
```

Result:

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

Query with exact Result:

```
SELECT
  CASE
      WHEN EXTRACT(HOUR FROM order_purchase_timestamp)
BETWEEN 0 AND 6 THEN 'Dawn'
      WHEN EXTRACT(HOUR FROM order_purchase_timestamp)
BETWEEN 7 AND 12 THEN 'Morning'
      WHEN EXTRACT(HOUR FROM order_purchase_timestamp)
BETWEEN 13 AND 18 THEN 'Afternoon'
      WHEN EXTRACT(HOUR FROM order_purchase_timestamp)
BETWEEN 19 AND 23 THEN 'Night'
  END AS time_of_day,
  COUNT(*) AS Number_of_orders
FROM target.orders
GROUP BY time_of_day
ORDER BY Number_of_orders DESC
LIMIT 1;
```

Result:

Row	time_of_day	Number_of_orders
1	Afternoon	38135

Insights and Recommendations:

Afternoon dominance:

- The query reveals that the highest number of orders are placed during the afternoon.
- This suggests that Brazilian customers are more active and engaged with online shopping during this time.
- Businesses should prioritize their operations and customer service during the afternoon to cater to the peak demand.

Potential marketing opportunities:

- With a significant number of orders being placed in the afternoon, businesses can leverage this information to optimize their marketing strategies.
- They can schedule targeted promotions, discounts, or advertisements during this time to attract more customers and drive sales.

Operational considerations:

- Since the afternoon is the busiest time for orders, businesses need to ensure their operational capacity can handle the increased demand.
- This includes having sufficient staff, efficient order processing systems, and streamlined logistics to fulfil orders promptly and maintain customer satisfaction.

Customer engagement:

- Analyzing the time of day when customers are most active provides an opportunity to engage with them effectively.
- Businesses can schedule email campaigns, social media posts, or live chat support during the afternoon to increase customer engagement and improve the overall shopping experience.

Personalization and recommendations:

- Based on the time of day, businesses can implement personalized recommendations to showcase relevant products or offers to customers.
- By analyzing past purchase patterns during the afternoon, businesses can identify popular items and recommend them to customers during that time, increasing the likelihood of sales.

Continuous monitoring:

- It's important to note that customer behavior and preferences can change over time. Therefore, businesses should continuously monitor the ordering patterns and regularly update their strategies accordingly.
- This could involve analyzing data over longer periods or specific seasonal trends to adapt marketing campaigns and operations effectively.

3.Evolution of E-commerce orders in the Brazil region:
a. Get the month on month no. of orders placed in each state.
Query:

```
SELECT
  c.customer_state,
     EXTRACT(YEAR FROM o.order_purchase_timestamp) AS purchase_year,
     EXTRACT(MONTH FROM o.order_purchase_timestamp) AS purchase_month,
  COUNT(*) AS Number_of_orders
FROM target.orders o
JOIN target.customers c ON o.customer_id = c.customer_id
GROUP BY c.customer_state, purchase_year, purchase_month
ORDER BY c.customer_state, purchase_year, purchase_month;
```

Result:

Row	customer_state	purchase_year	purchase_month	Number_of_orders
1	AC	2017	1	2
2	AC	2017	2	3
3	AC	2017	3	2
4	AC	2017	4	5
5	AC	2017	5	8
6	AC	2017	6	4
7	AC	2017	7	5
8	AC	2017	8	4
9	AC	2017	9	5
10	AC	2017	10	6

Insights and Recommendations:

Seasonal variations:

- Examining the monthly order count for each state can reveal seasonal trends in e-commerce orders.
- For example, state AC had a lower number of orders in January and February (2 and 3 orders, respectively) but experienced a significant increase in May (8 orders).
- This suggests a potential seasonal peak during that period. Identifying such patterns can help businesses allocate resources, plan inventory, and tailor marketing campaigns accordingly.

State-specific trends:

- Analyzing the data for each state individually can reveal unique order patterns.
- For instance, state AC had a gradual increase in orders from April to October, peaking in May (8 orders) and October (6 orders).
- Understanding state-specific trends can provide valuable insights into the factors influencing e-commerce activities and customer behavior in different regions.

Growth opportunities:

- States that exhibit consistent month-on-month growth in e-commerce orders indicate potential areas of opportunity for businesses.
- Monitoring states with a steady increase or higher order volumes can help identify regions where the e-commerce market is thriving.
- By focusing on these states, businesses can invest resources to expand their customer base, improve logistics, and tailor marketing efforts to capture the growing demand.

Customer behavior analysis:

- Associating customer states with order data allows for deeper analysis of customer behavior in different regions.
- By examining the patterns in each state, businesses can gain insights into preferences, purchasing power, and engagement levels of customers.
- This information can be used to create targeted marketing strategies and enhance the overall customer experience.

Regional marketing strategies:

- The month-on-month order count can help identify the best times to launch marketing campaigns in each state.
- For example, if a state consistently experiences a higher number of orders in a particular month, businesses can plan targeted promotions, discounts, or new product launches during that period to maximize customer engagement and sales.

Comparative analysis:

- Comparing the order counts between states can provide a better understanding of regional market dynamics.
- Identifying states with significantly higher or lower order volumes can help businesses benchmark their performance, discover untapped markets, or identify potential issues that need to be addressed.

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

Query:

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

Result:

Row	customer_state	Number_of_custome
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
11	PE	1652

Insights and Recommendations:

Concentration of customers:

- The query reveals that the state of São Paulo (SP) has the highest number of customers, with 41,746 unique customer IDs.
- This indicates a significant customer base concentrated in São Paulo.
- Businesses should consider this when developing marketing strategies, expanding operations, or targeting specific customer segments.

Regional customer distribution:

- The results highlight the distribution of customers across multiple states in Brazil.
- Along with São Paulo, states such as Rio de Janeiro (RJ), Minas Gerais (MG), Rio Grande do Sul (RS), and Paraná (PR) also have substantial customer populations.
- These states represent potential markets for businesses and may warrant additional attention when it comes to marketing campaigns or customer engagement initiatives.

Regional market potential:

- The distribution of customers across states can provide insights into the market potential and demand for e-commerce in different regions.
- States with a higher number of customers may indicate greater market opportunities and higher customer engagement.
- Businesses can leverage this information to prioritize marketing efforts and tailor strategies to tap into these markets effectively.

Customer acquisition focus:

- By analyzing the distribution of customers across states, businesses can identify areas with untapped potential.
- States with a relatively lower number of customers, such as Santa Catarina (SC), Bahia (BA), Distrito Federal (DF), Espírito Santo (ES), and Goiás (GO), may present opportunities for customer acquisition and market expansion.
- Targeted marketing campaigns and customer outreach efforts can help businesses attract new customers in these regions.

Regional preferences and cultural factors:

- The customer distribution across states can also provide insights into regional preferences and cultural factors that influence customer behaviour.
- Different states may have unique characteristics, preferences, or purchasing patterns. Understanding these regional nuances can help businesses tailor their products, marketing messages, and customer experience to better resonate with the specific customer base in each state.

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

a. 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.

Query:

```
WITH total_2017 AS (
SELECT SUM(p.payment_value) AS total_payment_2017
FROM target.orders o
JOIN target.payments p ON o.order_id = p.order_id
WHERE EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017
AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1
AND 8),
total_2018 AS
```

```
(
SELECT
SUM(p.payment_value) AS total_payment_2018
FROM target.orders o
JOIN target.payments p ON o.order_id = p.order_id
WHERE EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018
AND
EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1
AND 8
)
SELECT Round((total_payment_2018 - total_payment_2017) /
total_payment_2017 * 100,2) AS cost_increase_percentage
FROM total_2017, total_2018
```

Result:

Row	cost_increase_percentage
1	136.98

Insights and Recommendations:

- The provided query calculates the percentage increase in the cost of orders from the year 2017 to 2018, specifically for the months between January and August.
- The result indicates that there was a significant increase of approximately 136.98% in the cost of orders from 2017 to 2018, considering only the months between January and August.

Increased consumer spending:

- The substantial percentage increase in the cost of orders suggests a growth in consumer spending through e-commerce.
- It indicates that customers were willing to spend more on their online purchases during the specified period.

Expansion of e-commerce market:

- The significant increase in order costs implies that the e-commerce market experienced growth and expansion during that time.
- This growth could be attributed to factors such as increased internet penetration, improved trust in online transactions, and broader product availability.

Economic impact:

- The rise in e-commerce order costs indicates a positive impact on the economy.
- It signifies increased business revenue, job creation, and a boost to the overall economic activity associated with the e-commerce sector.

- This growth can have multiplier effects on various industries, including logistics, packaging, and digital marketing.

Market dynamics and competition:

- The observed increase in order costs may also reflect changes in market dynamics and competition.
- It suggests that businesses may have adjusted their pricing strategies to capture higher value transactions or to cover rising operational costs.

Consumer behaviour and purchasing power:

- The percentage increase in order costs highlights changes in consumer behavior and purchasing power.
- It suggests that customers were willing and able to spend more on their online purchases during the specified period, indicating increased consumer confidence and economic stability.

Business opportunities:

- The significant increase in order costs underscores the potential business opportunities within the e-commerce sector.
- It highlights the growing market demand and the potential for businesses to capitalize on this trend by offering high-value products, expanding product lines, or providing premium services.

Future projections:

- The observed increase in the cost of orders provides valuable insights for businesses and policymakers to make future projections and plan strategies accordingly.
- It indicates a positive trajectory for the e-commerce sector, prompting businesses to invest in infrastructure, technology, and customer experience to meet the evolving demands of online shoppers.

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

Query:

```
SELECT
  c.customer_state,
  ROUND(SUM(oi.price),2) AS total_order_price,
  Round(AVG(oi.price),2) AS average_order_price
FROM target.order_items oi
JOIN target.orders o ON oi.order_id = o.order_id
JOIN target.customers c ON o.customer_id = c.customer_id
GROUP BY c.customer_state;
```

Result:

Row	customer_state	total_order_price	average_order_price
1	MT	156453.53	148.3
2	MA	119648.22	145.2
3	AL	80314.81	180.89
4	SP	5202955.05	109.65
5	MG	1585308.03	120.75
6	PE	262788.03	145.51
7	RJ	1824092.67	125.12
8	DF	302603.94	125.77
9	RS	750304.02	120.34
10	SE	58920.85	153.04

Insights and Recommendations:

Regional variations:

- The total and average order prices vary across different states.
- For example, São Paulo (SP) has the highest total order price of 5,202,955.05, indicating a high overall sales volume.
- On the other hand, states like Sergipe (SE) and Alagoas (AL) have relatively lower total order prices, suggesting smaller market sizes or lower customer spending.

Consumer purchasing power:

- The average order price reflects the purchasing power of customers in each state.
- States with higher average order prices, such as Alagoas (AL) and Sergipe (SE), may indicate a higher willingness and ability of customers to spend more on their purchases.
- This information can be useful for businesses to tailor their product offerings and pricing strategies according to regional purchasing power.

Market potential:

- The total order price for each state provides insights into the market potential and the size of the e-commerce market in different regions.
- States with higher total order prices, like São Paulo (SP), Minas Gerais (MG), and Rio de Janeiro (RJ), represent larger market opportunities for businesses.
- These states may warrant additional focus in terms of marketing campaigns, customer acquisition strategies, and logistical operations.

Competitive landscape:

- The variations in total and average order prices across states also reflect the competitive landscape in different regions.
- Higher total order prices in certain states may indicate intense competition or the presence of premium products/services.
- Understanding the competitive dynamics can help businesses adjust their pricing, marketing, and customer retention strategies to remain competitive.

Business decision-making:

- The total and average order prices can aid businesses in making informed decisions.
- For instance, businesses can identify states with high average order prices to prioritize marketing efforts or explore opportunities for up-selling or cross-selling higher-priced products.
- On the other hand, states with lower average order prices may require different strategies, such as cost optimization or targeting specific customer segments.

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

Query:

```
SELECT
  c.customer_state,
  ROUND(SUM(oi.freight_value),2) AS total_freight_value,
  ROUND(AVG(oi.freight_value),2) AS average_freight_value
FROM target.order_items oi
JOIN target.orders o ON oi.order_id = o.order_id
JOIN target.customers c ON o.customer_id = c.customer_id
GROUP BY c.customer_state;
```

Result:

Row	customer_state	total_freight_value	average_freight_valu
1	SP	718723.07	15.15
2	RJ	305589.31	20.96
3	PR	117851.68	20.53
4	SC	89660.26	21.47
5	DF	50625.5	21.04
6	MG	270853.46	20.63
7	PA	38699.3	35.83
8	BA	100156.68	26.36
9	GO	53114.98	22.77
10	RS	135522.74	21.74

Insights and Recommendations:

Variations in freight costs:

- The total and average freight values differ across states. São Paulo (SP) has the highest total freight value of 718,723.07, indicating a significant volume of orders being shipped.
- Conversely, states like Pará (PA) and Goiás (GO) have lower total freight values, suggesting lower shipping volumes or potentially shorter average distances for deliveries.

Average freight costs:

- The average freight value reflects the typical cost of shipping an order in each state.
- States with higher average freight costs, such as Pará (PA) and Bahia (BA), may indicate longer distances or logistical challenges, resulting in higher shipping costs.
- On the other hand, states like São Paulo (SP) and Minas Gerais (MG) have relatively lower average freight costs, which may be attributed to their central geographical location and well-developed logistics networks.

Logistics considerations:

- The variations in total and average freight costs highlight the importance of logistics and shipping considerations in different states.
- Businesses operating in states with higher freight costs may need to optimize their supply chain, negotiate better shipping rates, or explore alternative delivery methods to maintain competitive pricing and customer satisfaction.

Customer expectations:

- The freight costs can influence customer expectations and purchasing decisions.
- Higher freight costs may impact customer willingness to place orders or may require businesses to absorb or offer competitive shipping rates.
- Understanding regional differences in freight costs can help businesses set appropriate pricing strategies and communicate shipping fees transparently to customers.

Market competitiveness:

- The variations in freight costs across states can indicate differences in market competitiveness.
- States with higher freight costs may face challenges in price competitiveness compared to states with lower freight costs.
- Businesses operating in states with higher freight costs need to factor these costs into their pricing strategies and explore ways to minimize their impact on overall profitability.

Strategic shipping decisions:

- The total and average freight costs provide insights for businesses to make strategic shipping decisions.
- For instance, businesses can identify states with higher average freight costs to focus on optimizing logistics, negotiating better rates with shipping providers, or leveraging economies of scale to reduce overall shipping expenses.

Customer experience:

- Freight costs directly impact the customer experience, as customers consider both product prices and shipping fees when making purchasing decisions.
- Businesses should aim to strike a balance between competitive pricing and reasonable shipping costs to provide a positive customer experience.

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

a. 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**

QUERY:

```
select order_id, date_diff(order_delivered_customer_date,
order_purchase_timestamp, day) as time_to_deliver,
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
order by order_id;
```

Result:

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

Insights and Recommendations:

Delivery Time:

- The analysis of delivery time (time_to_deliver) shows the number of days taken to deliver each order from the purchase date.
- The range of delivery time varies from a few days to several weeks.
- It is important for businesses to monitor and manage delivery times to ensure timely order fulfilment and meet customer expectations.

Estimated vs. Actual Delivery:

- The difference between the estimated and actual delivery date (diff_estimated_delivery) reveals the variation between the expected and actual delivery time.
- Positive values indicate delays in delivery, while negative values suggest early deliveries. This information can help businesses identify patterns and trends in their delivery performance and make necessary adjustments to improve accuracy and reliability.

Delivery Performance Discrepancies:

- By analyzing the delivery time and the difference between estimated and actual delivery, businesses can identify orders that experienced significant delays or were delivered earlier than expected.
- These discrepancies may be due to factors such as logistics issues, operational inefficiencies, or external factors affecting delivery timelines.
- Addressing these discrepancies is crucial to enhance customer satisfaction and optimize delivery processes.

Variations Across Orders:

- The query results highlight that delivery time and estimated delivery discrepancies can vary across different orders.
- Some orders are delivered within a short timeframe, while others experience longer delivery times.
- It is essential for businesses to evaluate the reasons behind these variations and implement strategies to minimize delays and ensure consistent delivery experiences for all customers.

Order Fulfillment Efficiency:

- Timely order delivery is crucial for customer satisfaction.
- By analyzing the data, businesses can identify orders with longer delivery times or significant discrepancies between estimated and actual delivery.
- This information can guide them in improving their order fulfillment efficiency, streamlining processes, and optimizing their logistics operations.

Customer Experience Management:

- Delivery time and accuracy are key factors that impact the overall customer experience. Businesses should focus on reducing delivery times and minimizing the discrepancies between estimated and actual delivery.
- Enhancing the customer experience through efficient and reliable delivery processes can lead to increased customer satisfaction, loyalty, and positive word-of-mouth recommendations.

Operational Optimization:

- Analyzing delivery metrics can provide insights into potential areas for operational improvement.
- By identifying trends and patterns in delivery times and discrepancies, businesses can take proactive measures to optimize their operations.
- This may involve reviewing logistics partnerships, streamlining order processing, improving inventory management, or enhancing communication with customers throughout the delivery process.

b.Find out the top 5 states with the highest & lowest average freight value.
Query:

```
SELECT
  customer_state,
  ROUND(AVG(freight_value),2) AS Highest_average_freight
FROM target.order_items
JOIN target.orders ON order_items.order_id = orders.order_id
JOIN target.customers ON orders.customer_id = customers.customer_id
GROUP BY customer_state
ORDER BY Highest_average_freight DESC
LIMIT 5; -- Highest average freight
```

Result:

Highest Average Freight

Row	customer_state	Highest_average_freight
1	RR	42.98
2	PB	42.72
3	RO	41.07
4	AC	40.07
5	PI	39.15

Query:

```
SELECT
  customer_state,
  ROUND(AVG(freight_value),2) AS Lowest_average_freight
FROM target.order_items
JOIN target.orders ON order_items.order_id = orders.order_id
JOIN target.customers ON orders.customer_id = customers.customer_id
GROUP BY customer_state
ORDER BY Lowest_average_freight
LIMIT 5; -- Lowest average freight
```

Result:

Lowest Average Freight

Row	customer_state	Lowest_average_freight
1	SP	15.15
2	PR	20.53
3	MG	20.63
4	RJ	20.96
5	DF	21.04

Insights and Recommendations:

Regional Variations:

- The average freight values vary significantly among different states in the dataset.
- This indicates that freight costs are influenced by regional factors such as distance, transportation infrastructure, and logistical challenges.

High Average Freight Values:

- States like Roraima (RR), Paraíba (PB), Rondônia (RO), Acre (AC), and Piauí (PI) have the highest average freight values.
- This could be due to factors like remote locations, limited transportation options, or higher logistics costs in these regions.

Low Average Freight Values:

- On the other hand, states like São Paulo (SP), Paraná (PR), Minas Gerais (MG), Rio de Janeiro (RJ), and Distrito Federal (DF) have the lowest average freight values.
- These states might benefit from better transportation infrastructure, more competitive logistics services, or shorter distances to major shipping hubs.

Cost Optimization Opportunities:

- Businesses operating in states with high average freight values should analyze their supply chain and logistics operations to identify cost optimization opportunities.
- This could involve negotiating better rates with shipping providers, exploring alternative transportation modes, or optimizing distribution networks.

Regional Pricing Strategies:

- The significant regional variations in average freight values suggest that businesses should consider implementing regional pricing strategies.
- Adjusting product prices based on freight costs in different states can help maintain profitability while remaining competitive in each region.

Supply Chain Efficiency:

- States with lower average freight values may indicate better supply chain efficiency, shorter lead times, or closer proximity to key markets.
- Businesses can leverage these insights to streamline their supply chain operations, improve inventory management, and enhance customer satisfaction by delivering products faster and at lower shipping costs.

Collaborative Shipping Initiatives:

- Businesses in states with high average freight values can collaborate with other companies to leverage collective bargaining power with shipping providers.
- By forming consortiums or alliances, companies can negotiate better rates, consolidate shipments, and optimize logistics operations.

Strategic Warehouse Placement:

- Analyzing the distribution patterns and average freight values across different states can help businesses strategically place warehouses and fulfillment centers.
- Locating facilities in states with lower average freight values can reduce shipping costs and improve order fulfillment efficiency.

C. Find out the top 5 states with the highest & lowest average delivery time.
Query:

```
SELECT t2.customer_state,
ROUND(AVG(TIMESTAMP_DIFF(order_delivered_customer_date,ord
er_purchase_timestamp,DAY)),2) AS time_to_delivery_avg,
from target.orders t1
join target.customers t2 on t1.customer_id=t2.customer_id
where t1.order_status='delivered'
group by t2.customer_state
order by time_to_delivery_avg DESC limit 5
```

Result:

Top 5 States with Highest average delivery time

Row	customer_state	time_to_delivery_avg
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32

Query:

```
SELECT t2.customer_state,
ROUND(AVG(TIMESTAMP_DIFF(order_delivered_customer_date,ord
er_purchase_timestamp,DAY)),2) AS time_to_delivery_avg,
from target.orders t1
join target.customers t2 on t1.customer_id=t2.customer_id
where t1.order_status='delivered'
group by t2.customer_state
ORDER BY time_to_delivery_avg LIMIT 5;
```

Result:

Lowest Average delivery time

Row	customer_state	time_to_delivery_avg
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48

Insights and Recommendations:

Regional Variations:

- The average delivery times vary significantly among different states in the dataset.
- This indicates that the efficiency and effectiveness of logistics and delivery services can vary based on the geographic location.

High Average Delivery Times:

- States like Roraima (RR), Amapá (AP), Amazonas (AM), Alagoas (AL), and Pará (PA) have the highest average delivery times.
- This could be due to factors such as remote locations, logistical challenges, or limited transportation infrastructure in these regions.

Low Average Delivery Times:

- On the other hand, states like São Paulo (SP), Paraná (PR), Minas Gerais (MG), Distrito Federal (DF), and Santa Catarina (SC) have the lowest average delivery times.
- These states may benefit from better logistics networks, efficient transportation systems, or closer proximity to distribution centers.

Operational Efficiency:

- States with lower average delivery times may indicate higher operational efficiency in logistics and delivery processes.
- Businesses can analyze the operations in these states to identify best practices, optimize processes, and improve delivery speed in regions with longer delivery times.

Customer Experience:

- Longer delivery times can negatively impact customer satisfaction.
- Businesses should closely monitor the delivery performance in states with higher average delivery times and take proactive measures to improve service quality and meet customer expectations.

Infrastructure and Transportation Considerations:

- States with higher average delivery times may require additional attention and investment in transportation infrastructure.
- Analyzing the specific challenges in these regions can help identify areas for improvement, such as enhancing road networks, expanding transportation options, or establishing fulfilment centres closer to customers.

Delivery Optimization Strategies:

- Businesses operating in states with higher average delivery times should focus on optimizing their delivery operations.
- This may involve implementing route optimization algorithms, enhancing last-mile delivery capabilities, or partnering with local delivery service providers to improve efficiency and reduce transit times.

Collaboration and Partnerships:

- Collaboration among businesses, logistics providers, and local authorities can help address the challenges faced in states with longer delivery times.
- By working together, stakeholders can identify and implement solutions that improve transportation infrastructure, streamline customs processes, and expedite deliveries.

Data-Driven Decision Making:

- Continuous monitoring and analysis of delivery performance, transit times, and customer feedback can provide valuable insights for decision making.
- Businesses should leverage data analytics to identify bottlenecks, measure performance metrics, and make data-driven decisions to optimize delivery processes and reduce average delivery times.

Customer Communication:

- Transparency and proactive communication with customers regarding delivery times can help manage expectations and improve customer satisfaction.
- Providing real-time tracking updates, delivery ETAs, and responsive customer support can mitigate the impact of longer delivery times and enhance the overall customer experience.

d. 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.

Query:

```
SELECT customer_state,
ROUND(AVG(date_diff(order_estimated_delivery_date,
order_purchase_timestamp,DAY)),2)AS delivery_time_difference
FROM target.orders
JOIN target.customers
ON orders.customer_id = customers.customer_id
WHERE order_delivered_customer_date IS NOT NULL
GROUP BY customer_state
ORDER BY delivery_time_difference DESC
LIMIT 5;
```

Result:

Row	customer_state	delivery_time_difference
1	AP	45.87
2	RR	45.63
3	AM	44.92
4	AC	40.72
5	RO	38.39

Insights and Recommendations:

Efficient Delivery Performance:

- The top 5 states identified (AP, RR, AM, AC, and RO) have consistently demonstrated faster delivery times compared to the estimated delivery dates.
- This suggests that logistics operations in these states are efficient and effective in delivering orders ahead of schedule.

Regional Efficiency:

- These states may have well-established transportation networks, optimized last-mile delivery processes, and effective coordination between logistics providers and retailers.
- This regional efficiency can contribute to faster order fulfilment and timely deliveries.

Customer Satisfaction:

- Faster delivery times compared to the estimated date of delivery can significantly enhance customer satisfaction.
- Customers in these states are likely to receive their orders earlier than expected, which can positively impact their overall experience and perception of retailer or e-commerce platform.

Process Optimization:

- While these states have shown strong delivery performance, continuous process optimization is crucial to maintain and improve efficiency.
- Regularly evaluate and optimize order fulfilment processes, such as inventory management, packaging, and transportation routing, to ensure timely and accurate deliveries.

Capacity Planning:

- As the demand for faster deliveries increases, businesses operating in these states should anticipate and plan for scalability and capacity expansion.
- This may involve investing in additional warehousing facilities, increasing delivery fleet size, or leveraging partnerships with third-party logistics providers to meet growing customer expectations.

Real-Time Tracking and Communication:

- Offer real-time order tracking and proactive communication with customers to keep them informed about their delivery status.
- Providing accurate and timely updates on delivery progress can further enhance customer satisfaction and minimize inquiries or concerns regarding the delivery process.

Data Analysis and Insights:

- Continuously analyze delivery performance data, including actual vs. estimated delivery times, to identify trends, patterns, and potential areas for improvement.
- Utilize data analytics tools to gain insights into operational efficiency, identify bottlenecks, and make data-driven decisions for process optimization.

Collaboration with Logistics Partners:

- Foster strong partnerships with logistics providers to ensure seamless coordination and streamline the delivery process.
- Collaborative efforts can help optimize transportation routes, leverage shared resources, and improve overall delivery performance.

6. Analysis based on the payments:

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

Query:

```
SELECT
  EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
    EXTRACT(MONTH FROM order_purchase_timestamp) AS order_month,
  payment_type,
  COUNT(*) AS num_orders
FROM target.orders
JOIN target.payments ON orders.order_id = payments.order_id
GROUP BY order_year, order_month, payment_type
ORDER BY order_year, order_month;
```

Result:

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

Insights and Recommendations:

Payment Type Distribution:

- The query provides information on the number of orders placed in different months categorized by payment types such as credit card, voucher, debit card, and UPI (Unified Payments Interface).
- This analysis allows us to understand the popularity and adoption of various payment methods over time.

Seasonal Variations:

- There are fluctuations in the number of orders placed with different payment types across months and years.

- This suggests that consumer behaviour and preferences might change throughout the year, potentially influenced by factors such as holidays, promotions, or economic factors.

Payment Type Preferences:

- By examining the data, we can identify the payment types that are more commonly used by customers.
- This information can help businesses understand customer preferences and tailor their payment options and strategies accordingly.

Diversify Payment Options:

- Based on the payment type distribution analysis, it is advisable for businesses to offer a variety of payment options to cater to different customer preferences.
- This may include popular methods such as credit cards, as well as emerging and convenient options like UPI or digital wallets.
- By providing multiple payment options, businesses can enhance customer convenience and capture a broader customer base.

Promote Secure Payment Methods:

- If credit card payments are prevalent, businesses should ensure robust security measures are in place to protect customer payment information.
- Additionally, educate customers about the security features of different payment methods to instil confidence and encourage the adoption of secure payment options.

Seasonal Marketing Strategies:

- Analyzing the seasonal variations in payment types can help businesses plan targeted marketing campaigns or promotions to align with customer preferences during specific months or events.
- For example, if UPI usage increases during certain months, businesses can design special offers or incentives to encourage customers to use this payment method.

Customer Payment Experience:

- Focus on enhancing the overall payment experience for customers.
- This includes providing a seamless checkout process, intuitive user interfaces, and clear instructions for various payment methods.
- A smooth payment experience can reduce cart abandonment rates and increase customer satisfaction.

Monitor Industry Trends:

- Keep a pulse on the evolving payment landscape and industry trends.
- Stay updated on emerging payment technologies, changing customer preferences, and regulatory developments.
- This will help businesses adapt their payment strategies and systems to meet evolving customer needs and stay competitive in the market.

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

Query:

```
SELECT
  payment_installments,
  COUNT(DISTINCT orders.order_id) AS num_orders
FROM target.payments
JOIN target.orders ON payments.order_id = orders.order_id
WHERE payment_installments > 0
GROUP BY payment_installments;
```

Result:

Row	payment_installment	num_orders
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

Insights and Recommendations:

Installment Payment Usage:

- Customers have utilized different payment installment options for their orders.
- The query provides information on the number of orders placed based on the number of installments chosen by customers.

Increasing Order Count with Fewer Installments:

- As the number of installments decreases, the number of orders placed tends to increase.
- This indicates that customers generally prefer to pay for their orders in fewer installments.

Flexible Payment Installment Options:

- Offer a range of installment options to cater to customers' diverse financial preferences.
- Providing flexibility in the number of installments can increase customer satisfaction and encourage more purchases.

Promote Lower Installment Options:

- Since there is a higher order count associated with fewer installments, businesses can promote and incentivize customers to choose lower installment plans.
- Offering discounts, rewards, or exclusive benefits for selecting a lower number of installments can encourage customers to opt for this payment method.

Clear Communication:

- Ensure that the installment payment options and their associated terms and conditions are clearly communicated to customers during the checkout process.
- Transparent and easy-to-understand information about installment plans can help customers make informed decisions and reduce any potential confusion or uncertainty.

Customer Insights:

- Further analyze customer behavior and preferences related to installment payments. Segment customers based on their payment installment choices and analyze their purchasing patterns, average order values, or product categories.
- This analysis can provide valuable insights for personalized marketing campaigns and targeted promotions.

Payment Processor Partnerships:

- Collaborate with payment processors or financial institutions to offer attractive installment payment options.
- Establishing partnerships can provide access to additional installment plans or financial services that can enhance the overall payment experience and attract more customers.

Through an emphasis on optimizing the online shopping experience, diversifying product offerings, and enhancing customer loyalty programs, Target can solidify its position as a leading retailer, attract new customers, and foster long-term growth in the ever-evolving market.