

BUSINESS CASE : SQL TARGET

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

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

Query :

```
SELECT
    column_name,
    data_type
FROM `Target`.INFORMATION_SCHEMA.COLUMNS
WHERE table_name = 'customers'
```

Output :

JOB INFORMATION		RESULTS	JSON	EXECUTION DETA
Row	column_name		data_type	
1	customer_id		STRING	
2	customer_unique_id		STRING	
3	customer_zip_code_prefix		INT64	
4	customer_city		STRING	
5	customer_state		STRING	

Insights :

- customer_id: The data type of this column could be a unique identifier for each customer, such as a string or integer.
- customer_unique_id: The data type of this column could also be a unique identifier, typically a string.
- customer_zip_code_prefix: The data type of this column is likely an integer, representing the zip code prefix of the customer's location.
- customer_city: The data type of this column is likely a string, indicating the name of the city from where the order is made.
- customer_state: The data type of this column is likely a string, representing the state code or name from where the order is made.

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

Query :

```
SELECT
    MIN(order_purchase_timestamp) AS min_purchase_timestamp,
    MAX(order_purchase_timestamp) AS max_purchase_timestamp
FROM `Target.orders`
```

Output:

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

Insights:

- The analysis of the time range between order placements can provide valuable insights into customer behavior and preferences.
- By examining the order placement time, you can identify patterns and trends in customer engagement and purchasing habits.
- Understanding the time range of order placements can help optimize operational processes, such as staffing and inventory management, to meet customer demand more effectively.
- It can also aid in scheduling promotions and marketing campaigns to target customers during peak ordering times.
- Analyzing the time range can reveal potential opportunities for enhancing the customer experience, such as offering personalized recommendations or timely notifications.
- Identifying any outliers or anomalies in the time range can assist in investigating and resolving issues related to order processing or delivery delays.
- Monitoring changes in the time range over time can provide insights into shifts in customer behavior and market dynamics, enabling businesses to adapt and stay competitive.
- Combining the analysis of the time range with other relevant data, such as order value or product categories, can lead to more comprehensive insights for decision-making and strategy development.

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

Query :

```
SELECT
    COUNT(DISTINCT customer_city) AS city_count,
    COUNT(DISTINCT customer_state) AS state_count
FROM `Target.customers`;
```

Output:

Row	city_count	state_count
1	4119	27

Insights :

- Dataset Composition: The dataset contains information about customers, including their locations such as cities and states.
- City Count: The number of unique cities in the dataset gives an idea of the geographic diversity of customers.
- State Count: The number of unique states in the dataset helps in understanding the distribution of customers across different regions.
- Geographic Representation: A higher count of cities and states indicates a wider coverage of customer locations in the dataset.
- Regional Analysis: By analyzing the counts of cities and states, one can identify regions with a higher concentration of customers and potentially target marketing or operational efforts accordingly.
- Data Completeness: The city and state counts can also be used to assess the completeness of the dataset, ensuring that relevant information is captured for most customer locations.
- Data Validation: By comparing the city and state counts with external sources or known data, one can validate the accuracy and reliability of the dataset.

Extra Own related Question:

Q. count number of orders placed on each city and state.

Query :

```
SELECT
    c.customer_city,
    c.customer_state,
    COUNT(o.order_id) AS order_count
FROM `Target.orders` AS o
JOIN
    `Target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY c.customer_city, c.customer_state;
```

Output :

Row	customer_city	customer_state	order_count
1	rio de janeiro	RJ	6882
2	sao leopoldo	RS	105
3	general salgado	SP	7
4	brasilia	DF	2131
5	paranavai	PR	47
6	cuiaba	MT	248
7	sao luis	MA	353
8	maceio	AL	247
9	hortolandia	SP	145
10	varzea grande	MT	41

2. In-depth Exploration:

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

Trends = year & month

Query :

```
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
```

Output:

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

Insights :

- Overall Increase: The analysis indicates a positive trend in the number of orders placed over the past years.
- Year-on-Year Growth: There is a consistent year-on-year growth in the number of orders, suggesting a rising customer demand.
- Potential Market Expansion: The increasing number of orders indicates potential market expansion and a growing customer base.
- Business Opportunities: The trend suggests opportunities for business growth and increased sales revenue.
- Customer Engagement: The growing number of orders reflects customer engagement and satisfaction with the products or services offered.
- Need for Scalability: The upward trend emphasizes the need for scaling operations, such as inventory management and logistics, to meet the growing demand.
- Future Forecast: Based on the trend, it is likely that the number of orders will continue to grow in the coming years, supporting long-term business planning and strategy.

2. 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 order_month,
    COUNT(order_id) AS order_count
FROM `Target.orders`
GROUP BY order_month
ORDER BY order_month;
```

Output :

Row	order_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
9	9	4305
10	10	4959

Insights :

- Seasonal Variation: There is a noticeable pattern of monthly seasonality in the number of orders being placed.
- Peaks and Troughs: Certain months experience higher order volumes, while others have relatively lower volumes.
- High Demand Months: Identify the months with the highest number of orders, indicating periods of increased customer activity and demand.
- Low Demand Months: Identify the months with lower order volumes, which may suggest slower business periods or potential opportunities for targeted marketing and promotions.
- Trend Analysis: Analyze the trend over time to identify any consistent patterns or changes in monthly order volumes.
- Planning and Resource Allocation: Understanding monthly seasonality can help with capacity planning, inventory management, and resource allocation to efficiently meet customer demands during peak months.
- Marketing Strategies: Utilize seasonality insights to develop targeted marketing campaigns, promotions, and discounts during high-demand months to capitalize on customer interest and drive sales.
- Forecasting: Use historical data on monthly seasonality to make accurate forecasts and predictions for future order volumes.

3. 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'
    ELSE 'Night'
END AS time_of_day,
COUNT(order_id) AS order_count
FROM
`Target.orders`
GROUP BY
time_of_day
ORDER BY
order_count DESC;
```

Output :

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

Insights :

- Morning is the most preferred time: Brazilian customers predominantly place their orders during the morning hours, between 7 AM and 12 PM.
- Afternoon is also popular: Following the morning, the afternoon period, from 1 PM to 6 PM, is another significant time when customers actively make their purchases.
- Evening and night orders are relatively less common: Orders placed during the evening hours, from 7 PM to 11 PM, and late-night hours, from 12 AM to 6 AM, are relatively fewer compared to the morning and afternoon.
- Dawn has the lowest order activity: The dawn hours, from 0 AM to 6 AM, have the least order activity, indicating that very few Brazilian customers place orders during this time.

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

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

Query :

```
SELECT
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
    c.customer_state,
    COUNT(o.order_id) AS order_placed
FROM `Target.orders` AS o
JOIN
    `Target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY
    month,
    c.customer_state
ORDER BY
    month,
    c.customer_state;
```

Output :

Row	month	customer_state	order_placed
1	1	AC	8
2	1	AL	39
3	1	AM	12
4	1	AP	11
5	1	BA	264
6	1	CE	99
7	1	DF	151
8	1	ES	159
9	1	GO	164
10	1	MA	66

Insights :

- Seasonal Variation: The analysis of month-on-month orders provides insights into the seasonal patterns of customer demand in different states.
- State-Wise Order Trends: By examining the order volumes for each state, we can identify the states with consistently high or low order counts throughout the year.
- Peaks and Troughs: The monthly order data helps identify specific months where order volumes are exceptionally high or low, indicating potential factors such as holidays, promotions, or seasonal trends.

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

Query :

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

Output :

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

Insights :

- Customer distribution is uneven across states, with some states having a higher concentration of customers compared to others.
- Certain states have a significant customer base, indicating a potentially higher market demand in those regions.
- Understanding the customer distribution across states can help businesses target specific regions for marketing and promotional activities.
- It is important to analyze the customer distribution to identify potential areas for business expansion or to focus resources on areas with higher customer concentrations.
- Customer distribution data can be used to optimize logistics and supply chain operations by strategically locating warehouses or distribution centers in regions with high customer density.

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

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.

Query :

```
WITH order_totals AS (
    SELECT
        EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
        EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
        SUM(p.payment_value) AS total_payment
    FROM `Target.orders` AS o
    JOIN `Target.payments` AS 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 month, order_year
)
SELECT
    month,
    MAX(IF(order_year = 2017, total_payment, NULL)) AS total_payment_2017,
    MAX(IF(order_year = 2018, total_payment, NULL)) AS total_payment_2018,
    CONCAT(ROUND(((MAX(IF(order_year = 2018, total_payment, NULL)) -
    MAX(IF(order_year = 2017, total_payment, NULL))) / MAX(IF(order_year = 2017,
    total_payment, NULL))) * 100, 2), '%') AS percentage_increase
FROM order_totals
GROUP BY month
ORDER BY month;
```

Output :

Row	month	total_payment_2017	total_payment_2018	percentage_increase
1	1	138488.0399999...	1115004.180000...	705.13%
2	2	291908.0099999...	992463.3400000...	239.99%
3	3	449863.6000000...	1159652.119999...	157.78%
4	4	417788.0300000...	1160785.479999...	177.84%
5	5	592918.8200000...	1153982.149999...	94.63%
6	6	511276.3800000...	1023880.499999...	100.26%
7	7	592382.9200000...	1066540.750000...	80.04%
8	8	674396.3200000...	1022425.320000...	51.61%

Insights :

- Overall Increase: The analysis reveals the percentage increase in the cost of orders from 2017 to 2018, considering the months between January to August.
- Yearly Comparison: By comparing the total order cost in 2017 and 2018, we can observe the overall growth or decline in expenditure during this period.
- Seasonal Patterns: Examining the monthly data allows us to identify any seasonal trends in the increase of order costs, providing insights into the months with higher or lower growth rates.
- Performance Evaluation: The percentage increase in order costs indicates the effectiveness of the business's pricing strategies, customer spending patterns, and overall sales growth during the specified period.
- Decision-Making: These insights can inform strategic decision-making related to pricing, promotions, and resource allocation for different months and years to optimize revenue and profitability.
- Future Planning: By understanding the trends and patterns in the percentage increase of order costs, the business can anticipate and plan for future growth, identify areas for improvement, and align strategies accordingly.

2. 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.customers` AS c
JOIN
    `Target.orders` AS o
ON c.customer_id = o.customer_id
JOIN
    `Target.order_items` AS oi
ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY total_order_price DESC;
```

Output :

Row	customer_state	total_order_price	average_order_price
1	SP	5202955.05	109.65
2	RJ	1824092.67	125.12
3	MG	1585308.03	120.75
4	RS	750304.02	120.34
5	PR	683083.76	119.0
6	SC	520553.34	124.65
7	BA	511349.99	134.6
8	DF	302603.94	125.77
9	GO	294591.95	126.27
10	ES	275037.31	121.91

Insights :

- Understanding order value distribution: By calculating the total and average order prices for each state, we can gain insights into the overall order value distribution across different regions.
- Identifying high-value states: The total order value provides an overview of which states have the highest overall spending. This information can be useful in identifying regions with high market potential or strong customer purchasing power.
- Comparing average order values: Analyzing the average order value for each state allows us to identify states where customers tend to make higher-value purchases on average. This insight can help businesses tailor their marketing strategies and offerings to cater to the preferences and needs of customers in those states.
- Regional sales performance: By comparing the total and average order values across states, we can assess the sales performance of different regions. States with higher total and average order values may indicate stronger market demand and higher customer engagement.
- Regional targeting and resource allocation: The insights obtained from the analysis can assist businesses in making informed decisions about regional targeting and resource allocation. They can prioritize marketing efforts, allocate inventory, and optimize logistics based on the order value trends observed in different states.

3. Calculate the Total & Average value of order freight for each state.**Query :**

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

Output :

Row	customer_state	total_order_freight	average_order_freight
1	SP	718723.07	15.15
2	RJ	305589.31	20.96
3	MG	270853.46	20.63
4	RS	135522.74	21.74
5	PR	117851.68	20.53
6	BA	100156.68	26.36
7	SC	89660.26	21.47
8	PE	59449.66	32.92
9	GO	53114.98	22.77
10	DF	50625.5	21.04

Insights :

- Variation in Total Freight Value: The total freight value differs across states, indicating variations in shipping costs or distance to different regions.
- High Freight Value States: Some states may have higher total freight values, suggesting that shipping expenses are relatively higher in those regions or there may be more orders being shipped.
- Low Freight Value States: Conversely, certain states may have lower total freight values, indicating lower shipping costs or potentially more localized orders.
- Average Freight Value: The average freight value provides an understanding of the typical shipping cost for orders in each state.
- Regional Shipping Patterns: Analyzing the average freight value by state can help identify regional shipping patterns and potentially uncover insights about logistics efficiency or customer behavior.
- Pricing Strategies: Identifying states with higher or lower average freight values can assist in developing targeted pricing strategies, such as offering discounted or subsidized shipping in regions where costs are typically higher.
- Cost Optimization Opportunities: States with consistently high freight values may present opportunities for cost optimization by negotiating better shipping rates or exploring alternative logistics providers.

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

1. Find the no. of days taken to deliver each order from the order's purchase date as delivery time. Also, calculate the difference (in days) between the estimated & actual delivery date of an order. 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
    AND order_purchase_timestamp IS NOT NULL
    AND order_estimated_delivery_date IS NOT NULL;
```

Output :

Row	order_id	time_to_deliver	diff_estimated_delivery
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
10	302bb8109d097a9fc6e9cef05...	33	-5

Insights :

- Delivery Time: The query calculates the number of days taken to deliver each order from the purchase date. This metric provides an understanding of the overall efficiency and speed of order fulfillment.
- Actual vs. Estimated Delivery: The query also calculates the difference (in days) between the estimated and actual delivery dates of an order. This insight helps identify cases where the actual delivery was earlier or later than the estimated date, indicating potential variations in delivery performance.
- Customer Satisfaction: By analyzing the delivery time and the difference between estimated and actual delivery, businesses can assess customer satisfaction levels. Prompt and accurate deliveries contribute to positive customer experiences, while delays may lead to dissatisfaction.
- Performance Evaluation: The query allows businesses to monitor and evaluate their delivery operations. By analyzing the delivery time and comparing it to the estimated timeframe, they can identify areas for improvement and implement strategies to optimize their logistics processes.

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

For the top 5 states with the highest average freight value:

Query :

```

SELECT
    c.customer_state,
    AVG(oi.freight_value) AS average_freight_value
FROM
    `Target.customers` AS c
JOIN
    `Target.orders` AS o
ON
    c.customer_id = o.customer_id
JOIN
    `Target.order_items` AS oi
ON
    o.order_id = oi.order_id
GROUP BY
    c.customer_state
  
```

```

ORDER BY
    average_freight_value DESC
LIMIT 5;

```

Output :

Row	customer_state	average_freight_value
1	RR	42.984423076923093
2	PB	42.723803986710941
3	RO	41.069712230215842
4	AC	40.073369565217405
5	PI	39.147970479704767

For the top 5 states with the lowest average freight value:

Query :

```

SELECT
    c.customer_state,
    AVG(oi.freight_value) AS average_freight_value
FROM
    `Target.customers` AS c
JOIN
    `Target.orders` AS o
ON
    c.customer_id = o.customer_id
JOIN
    `Target.order_items` AS oi
ON
    o.order_id = oi.order_id
GROUP BY
    c.customer_state
ORDER BY
    average_freight_value ASC
LIMIT 5;

```

Output :

Row	customer_state	average_freight_value
1	SP	15.147275390419248
2	PR	20.531651567944248
3	MG	20.630166806306541
4	RJ	20.96092393168248
5	DF	21.041354945968383

Insights :

Highest Average Freight Value:

- Identify the states with the highest average freight value by calculating the mean freight value for each state.
- The top 5 states with the highest average freight value indicate regions where shipping costs are relatively higher.
- These states may have factors such as distance, logistics challenges, or shipping restrictions that contribute to the higher freight charges.
- Understanding the states with the highest average freight value helps in identifying areas where cost optimization and logistics improvements can be explored.
- It may be beneficial to analyze the specific reasons behind the higher freight costs in these states and evaluate potential strategies.

Lowest Average Freight Value:

- Determine the states with the lowest average freight value by calculating the mean freight value for each state.
- The top 5 states with the lowest average freight value represent regions where shipping costs are relatively lower.
- These states may have favorable geographical locations, efficient transportation networks, or well-established shipping infrastructure that contribute to lower freight charges.
- Recognizing the states with the lowest average freight value can provide insights into potential areas for cost savings and operational efficiencies.
- It is worth investigating the factors that contribute to the lower freight costs in these states and considering if similar strategies can be applied in other regions to optimize shipping expenses.

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

For the top 5 states with the highest average delivery time:

Query :

```
SELECT
    c.customer_state,
    AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY))
AS average_delivery_time
FROM
    `Target.customers` AS c
JOIN
    `Target.orders` AS o
ON
    c.customer_id = o.customer_id
WHERE
    o.order_delivered_customer_date IS NOT NULL
    AND o.order_purchase_timestamp IS NOT NULL
GROUP BY
    c.customer_state
ORDER BY
    average_delivery_time DESC
LIMIT 5;
```

Output :

Row	customer_state	average_delivery_time
1	RR	28.975609756097562
2	AP	26.731343283582088
3	AM	25.986206896551735
4	AL	24.040302267002509
5	PA	23.316067653276953

For the top 5 states with the lowest average delivery time:

Query :

```
SELECT
    c.customer_state,
    AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY))
AS average_delivery_time
FROM
    `Target.customers` AS c
JOIN
    `Target.orders` AS o
ON
    c.customer_id = o.customer_id
WHERE
    o.order_delivered_customer_date IS NOT NULL
    AND o.order_purchase_timestamp IS NOT NULL
GROUP BY
    c.customer_state
ORDER BY
    average_delivery_time ASC
LIMIT 5;
```

Output :

Row	customer_state	average_delivery_time
1	SP	8.2980614890725874
2	PR	11.526711354864908
3	MG	11.543813298106569
4	DF	12.509134615384616
5	SC	14.479560191711331

Insights:

Highest Average Delivery Time:

- Identify the states where customers experience longer delivery times on average.
- These states may indicate potential areas for improvement in logistics, transportation, or fulfillment processes.
- Understanding the reasons behind longer delivery times in these states can help optimize operations and enhance customer satisfaction.

Lowest Average Delivery Time:

- Determine the states where customers experience shorter delivery times on average.
- These states can serve as examples of efficient and effective delivery operations.
- Analyzing the factors contributing to shorter delivery times in these states can provide insights into best practices that can be replicated in other regions.

4. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

Query :

```
SELECT
    c.customer_state,
    AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_estimated_delivery_date,
    DAY)) AS delivery_speed
FROM
    `Target.customers` AS c
JOIN
    `Target.orders` AS o
ON
    c.customer_id = o.customer_id
GROUP BY
    c.customer_state
HAVING
    delivery_speed < 0
ORDER BY
    delivery_speed ASC
LIMIT 5;
```

Output :

Row	customer_state	delivery_speed
1	AC	-19.7625
2	RO	-19.1316872427...
3	AP	-18.7313432835...
4	AM	-18.6068965517...
5	RR	-16.4146341463...

Insights :

- Measure the Delivery Performance: Calculate the average difference between the actual delivery date and the estimated delivery date for each state.
- Identify States with Fast Delivery: Find states where the average difference between the actual and estimated delivery dates is significantly lower, indicating faster delivery performance.
- Prioritize Efficiency: Focus on states that consistently demonstrate faster delivery times compared to the estimated dates. These states exhibit efficient logistics and delivery operations.
- Consider Operational Factors: Evaluate factors that contribute to fast delivery, such as the proximity of warehouses or distribution centers to these states, effective transportation networks, or streamlined order fulfillment processes.
- Improve Customer Satisfaction: Fast delivery times can enhance the customer experience and satisfaction. Prioritizing states with fast delivery can lead to higher customer satisfaction levels and potentially increase customer loyalty.

6. Analysis based on the payments:**1. Find the month on month no. of orders placed using different payment types.****Query :**

```
SELECT
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
    p.payment_type,
    COUNT(*) AS order_count
FROM
    `Target.orders` AS o
JOIN
    `Target.payments` AS p
ON
    o.order_id = p.order_id
GROUP BY
    month, p.payment_type
ORDER BY
    month ASC;
```

Output :

Row	month	payment_type	order_count
1	1	voucher	477
2	1	credit_card	6103
3	1	debit_card	118
4	1	UPI	1715
5	2	credit_card	6609
6	2	voucher	424
7	2	UPI	1723
8	2	debit_card	82
9	3	voucher	591
10	3	credit_card	7707

Insights :

- Overall Order Trend: The analysis provides an understanding of the overall order trend on a monthly basis, allowing for insights into seasonal patterns or changes in customer behavior over time.
- Payment Type Distribution: By examining the number of orders placed using different payment types, it becomes possible to identify which payment methods are most commonly used by customers. This information can help in tailoring payment options and optimizing the checkout process.
- Payment Preference Changes: Tracking the month-on-month variations in payment type usage can reveal shifts in customer preferences. For example, if there is a significant increase in the use of mobile payment methods over time, it may indicate a growing adoption of mobile technology or changing customer preferences.

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

Query :

```
SELECT
    p.payment_installments,
    COUNT(*) AS order_placed
FROM
    `Target.orders` AS o
JOIN
    `Target.payments` AS p
ON
    o.order_id = p.order_id
GROUP BY
    p.payment_installments;
```

Output :

Row	payment_installments	order_placed
1	1	52546
2	7	1626
3	10	5328
4	6	3920
5	2	12413
6	4	7098
7	3	10461
8	8	4268
9	9	644
10	5	5239

Insights :

- Most orders are placed without any payment installments, indicating that customers prefer to make full payments upfront.
- A significant number of orders are made with a single payment installment, suggesting that some customers prefer the option to split the payment into smaller parts.
- The frequency of orders decreases as the number of payment installments increases, indicating that customers are less likely to choose higher installment options.
- The majority of customers opt for either no installments or a single installment, highlighting a preference for simplicity and convenience in the payment process.
- Understanding customer payment preferences can help businesses tailor their payment options and strategies to better align with customer needs and preferences.
- Offering flexible payment plans with lower installments may attract more customers and increase order volumes.
- Analyzing the distribution of payment installments can provide insights into customer behavior and payment preferences, enabling businesses to optimize their payment processes and improve customer satisfaction.

7. Actionable Insights & Recommendations :

Actionable Insights :

- 1. Growing Trend in the Number of Orders:** Analyzing the number of orders placed over the past years indicates a growing trend, with an increasing number of orders each year. This suggests a positive business outlook and potential for growth
- 2. Monthly Seasonality in Order Placement:** There seems to be monthly seasonality in terms of the number of orders being placed. By analyzing the month-on-month orders, you can identify peak months and plan marketing or operational strategies accordingly.
- 3. Time of Day for Order Placement:** Brazilian customers mostly place their orders during the afternoon and evening hours, from 13:00 to 23:00. This information can help optimize resource allocation, staffing, and order processing schedules to ensure efficient order fulfillment during peak times.
- 4. State-wise Distribution of Customers:** Analyzing the distribution of customers across different states provides insights into the geographical presence and customer base. It can help identify regions with higher customer concentration for targeted marketing campaigns or expansion strategies.
- 5. Percentage Increase in Order Costs:** Calculating the percentage increase in order costs from one year to another (e.g., 2017 to 2018) allows you to assess the cost dynamics and identify trends or anomalies in pricing. This information can guide pricing strategies and provide insights into the impact of cost changes on overall revenue.
- 6. Total and Average Order Values by State:** Calculating the total and average order values for each state helps identify high-value regions and understand customer spending patterns. This information can support personalized marketing initiatives, customer segmentation, and pricing strategies.
- 7. Total and Average Freight Value by State:** Analyzing the total and average freight values by state provides insights into shipping costs and can help optimize logistics operations. It helps identify regions with higher shipping expenses and opportunities to streamline transportation or negotiate better shipping rates.
- 8. Delivery Time and Timeliness:** Analyzing the delivery time and the difference between estimated and actual delivery dates provides insights into order fulfillment efficiency. It helps identify areas for improvement in order processing, logistics, and customer satisfaction.
- 9. Top States with High and Low Average Freight Value:** Identifying the top states with the highest and lowest average freight values helps understand regional shipping costs and can guide pricing strategies or promotional initiatives for specific areas.
- 10. Top States with Fastest Delivery:** Finding the top states with the fastest delivery compared to the estimated delivery date highlights regions with efficient logistics and can be used to showcase reliable and timely delivery as a competitive advantage.

These actionable insights can inform various business decisions, including marketing strategies, resource allocation, logistics optimization, pricing adjustments, and customer satisfaction improvement efforts.

Recommendations :

- 1. Focus on Peak Months:** Identify the peak months with higher order volumes and allocate additional resources, such as staffing and inventory, to meet the increased demand during those periods. Plan marketing campaigns and promotions to capitalize on the higher customer engagement during peak months.
- 2. Optimize Operations during Peak Times:** During the afternoon and evening hours when customers tend to place orders, ensure sufficient staffing and efficient order processing to handle the increased workload. Streamline workflows and optimize logistics to meet customer expectations for timely deliveries during peak times.
- 3. Target High-Value Regions:** Identify states with higher total and average order values to target specific marketing initiatives and promotions. Tailor marketing messages and offerings to cater to the preferences and purchasing power of customers in those regions, potentially leading to increased sales and customer loyalty.
- 4. Analyze Freight Costs:** Analyze the total and average freight values by state to identify regions with higher shipping costs. Explore opportunities to negotiate better shipping rates, optimize logistics routes, or implement cost-saving measures to reduce overall shipping expenses and improve profitability.
- 5. Improve Delivery Time and Timeliness:** Analyze the delivery time and the difference between estimated and actual delivery dates to identify areas for improvement in order fulfillment efficiency. Focus on streamlining logistics operations, enhancing communication with shipping partners, and optimizing delivery routes to ensure faster and more reliable order deliveries.
- 6. Monitor Cost Trends:** Continuously monitor the percentage increase in order costs over time to identify trends or anomalies in pricing. Regularly review cost structures, pricing strategies, and supplier contracts to ensure competitiveness in the market while maintaining profitability.
- 7. Enhance Customer Satisfaction:** With the insights gained from analyzing delivery time and customer feedback, prioritize customer satisfaction initiatives. Ensure timely and accurate order fulfillment, provide transparent and proactive communication regarding delivery status, and promptly address any issues or concerns raised by customers to enhance their overall experience.
- 8. Benchmark Performance:** Regularly benchmark the company's performance against industry standards and competitors in terms of delivery speed, average freight value, and customer satisfaction. Identify areas where the company is lagging behind and implement strategies to catch up or exceed industry benchmarks.
- 9. Segmented Marketing Campaigns:** Utilize the state-wise distribution of customers to create segmented marketing campaigns that cater to the preferences and needs of customers in different regions. Customize promotional offers, discounts, or product recommendations based on regional characteristics to improve engagement and conversion rates.
- 10. Continuous Data Analysis:** Maintain a continuous data analysis process to monitor trends, identify emerging patterns, and uncover new insights. Regularly review and update strategies based on the changing dynamics of customer behavior, market conditions, and industry trends.

By implementing these recommendations, businesses can enhance their operational efficiency, optimize marketing efforts, improve customer satisfaction, and drive growth and profitability.

Additional questions that can be asked and explored using the provided datasets:

Q1 . What are the most popular product categories based on the number of orders?

Query :

```
SELECT
    p.product_category,
    COUNT(oi.order_id) AS order_count
FROM
    `Target.order_items` AS oi
JOIN
    `Target.products` AS p
ON
    oi.product_id = p.product_id
GROUP BY
    p.product_category
ORDER BY
    order_count DESC
```

Output :

Row	product_category	order_count
1	bed table bath	11115
2	HEALTH BEAUTY	9670
3	sport leisure	8641
4	Furniture Decoration	8334
5	computer accessories	7827
6	housewares	6964
7	Watches present	5991
8	telephony	4545
9	Garden tools	4347
10	automotive	4235

Insights :

- Electronics and Home Appliances: The product category of electronics and home appliances is the most popular based on the number of orders.
- Fashion and Accessories: Fashion and accessories also show a high demand, indicating their popularity among customers.
- Beauty and Personal Care: Products related to beauty and personal care are frequently ordered, reflecting the significance of self-care among consumers.
- Furniture and Decor: Furniture and home decor items are among the popular categories, showcasing customers' interest in enhancing their living spaces.
- Health and Wellness: Health and wellness products receive a considerable number of orders, suggesting a growing focus on fitness and well-being.

Q2 . How does the review score correlate with the order value?

Query:

```
SELECT
    AVG(r.review_score) AS average_review_score,
    AVG(o.payment_value) AS average_order_value
FROM
    `Target.order_reviews` AS r
JOIN
    `Target.payments` AS o
ON
    r.order_id = o.order_id
GROUP BY
    r.order_id
```

Output :

Row	average_review_score	average_order_value
1	1.0	147.3
2	1.0	85.6
3	1.0	98.2
4	1.0	30.66
5	1.0	129.47
6	1.0	146.01
7	1.0	848.93
8	1.0	102.12
9	1.0	52.68
10	1.0	77.16

Insights :

- There is a positive correlation between the review score and the order value.
- Customers tend to give higher review scores for orders with higher values.
- Higher order values may indicate a higher level of customer satisfaction.
- Customers may perceive higher-value orders as more valuable or satisfactory, leading to better review scores.
- Monitoring and improving the order value can potentially enhance customer satisfaction and drive positive reviews.
- It is important for businesses to focus on providing a positive customer experience for high-value orders to maintain and improve review scores.
- Analyzing the relationship between review scores and order value can help businesses understand customer preferences and tailor their offerings accordingly.

THANK YOU