

Q1. 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 :

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EX
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:

An insightful observation gleaned from the data is the comprehensive exploration of data types across all columns within the Customers table, residing in the Target dataset.

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

Query :

```
SELECT
    MIN(order_purchase_timestamp) as min_,
    MAX(order_purchase_timestamp) as max_
FROM
    target.orders;
```

Output :

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECU
Row	min_ ▾	max_ ▾			
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC			

Insights :

We've uncovered the starting moment when the first order was made, along with the moment when the last order was placed. This sheds light on the entire timeline of order placements within the dataset.

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

Query :

```
with state_cte as
(
SELECT
customer_state,
COUNT(customer_id) as customer_state_count
FROM
`target.customers`
GROUP BY
customer_state
ORDER BY
customer_state
),
city_cte as
(
SELECT
customer_state,
customer_city,
COUNT(customer_id) as customer_city_count
FROM
```

```

        `target.customers`
GROUP BY
    customer_city,
    customer_state
ORDER BY
    customer_state,
    customer_city
)
SELECT
    c1.customer_state,
    c1.customer_state_count,
    c2.customer_city,
    c2.customer_city_count
FROM
    state_cte c1
    JOIN
    city_cte c2 USING(customer_state)
ORDER BY
    c1.customer_state,
    c2.customer_city
LIMIT
    10
;

```

Output :

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAP
Row	customer_state	customer_state_count	customer_city	customer_city_count		
1	AC	81	brasileia	1		
2	AC	81	cruzeiro do sul	3		
3	AC	81	epitaciolandia	1		
4	AC	81	manoel urbano	1		
5	AC	81	porto acre	1		
6	AC	81	rio branco	70		
7	AC	81	senador guiomard	2		
8	AC	81	xapuri	2		
9	AL	413	agua branca	1		
10	AL	413	anadia	2		

Insights :

An insightful observation from the data showcases the distribution of customer counts across states and cities. For instance, the state 'AC' has a total customer count of 81, with one of its cities, 'Brasileia', having 1 customer. This insight provides valuable insights into the geographic concentration of customers within states and their respective cities.

Recommendation :

Localized Product Offerings: Customize product offerings and assortments based on the preferences and needs of customers in different states and cities. Analyze purchase patterns and demand trends to identify product categories or items that are popular in specific regions.

Enhanced Customer Service: Prioritize customer service efforts in regions with high customer concentrations to deliver exceptional experiences and foster loyalty. Invest in customer support resources, such as dedicated representatives or localized customer service centers, to provide personalized assistance and resolve inquiries promptly.

Q2. In-depth Exploration:

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

Query :

```
SELECT
    Extract(YEAR from order_purchase_timestamp) as Year,
    COUNT(order_id) as order_count_over_years
FROM
    `target.orders`
GROUP BY
    Extract(YEAR from order_purchase_timestamp)
ORDER BY
    Year asc;
```

Output :

Query results

JOB INFORMATION		RESULTS	CHART
Row	Year ▼	order_count_over_years	
1	2016	329	
2	2017	45101	
3	2018	54011	

Insights :

When analyzing the trend of order placements over the past few years, it's evident that there was a significant surge in orders between 2016 and 2017, with a notable increase of around 45,000 orders. Subsequently, between 2017 and 2018, although the growth slowed down, there was still a steady rise, marked by an increase of approximately 9,000 orders.

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

Query :

```
with month_seasonality as
(
SELECT
  Extract(YEAR from order_purchase_timestamp) as Year,
  Extract(MONTH from order_purchase_timestamp) as Month,
  COUNT(order_id) as order_count_over_years_months,
FROM
  `target.orders`
GROUP BY
  Extract(YEAR from order_purchase_timestamp),
  Extract(MONTH from order_purchase_timestamp)
)
SELECT
  Year,
  Month,
  order_count_over_years_months
FROM
  month_seasonality
ORDER BY
  Month,
  Year;
```

Output :

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXEC
Row	Year ▼	Month ▼	order_count_over_ye		
1	2017	Month	1	800	
2	2018		1	7269	
3	2017		2	1780	
4	2018		2	6728	
5	2017		3	2682	
6	2018		3	7211	
7	2017		4	2404	
8	2018		4	6939	
9	2017		5	3700	
10	2018		5	6873	
11	2017		6	3245	

Insights :

From the dataset provided, it's apparent that there's a distinct pattern in order placements. Specifically, between April and August, there's a notable spike in the number of orders placed by customers compared to other months. This suggests a seasonal trend, indicating that certain times of the year are busier for customer orders.

Recommendation :

Seasonal Promotions: Capitalize on the increased order activity during peak months (April to August) by launching seasonal promotions and marketing campaigns. Offer special discounts, bundles, or limited-time offers tailored to coincide with high-demand periods. This can incentivize customers to make purchases during these busy months, thereby boosting revenue.

Inventory Management: Anticipate higher order volumes during peak months and adjust inventory levels accordingly. Ensure sufficient stock of popular items to meet increased demand, while also avoiding overstocking to minimize storage costs and reduce the risk of inventory obsolescence. Utilize data analytics and forecasting tools to optimize inventory management strategies.

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
    t2.time_,
    COUNT(t2.time_) as order_count
FROM
    (
        SELECT
            t1.order_id,
            CASE
                WHEN t1.Hour BETWEEN 0 AND 6 THEN 'Dawn'
                WHEN t1.Hour BETWEEN 7 AND 12 THEN 'Mornings'
                WHEN t1.Hour BETWEEN 13 AND 18 THEN 'Afternoon'
                WHEN t1.Hour BETWEEN 19 AND 23 THEN 'Night'
            END as time_
        FROM
            (
                SELECT
                    *,
                    Extract(hour from order_purchase_timestamp) as Hour
                FROM
                    `target.orders`
            )t1
        )t2
GROUP BY
    T2.time_
ORDER BY
    order_count desc;
```


Output :

Query results

JOB INFORMATION		RESULTS	CHART	JSON
Row	time_ ▾	order_count ▾		
1	Afternoon	38135		
2	Night	28331		
3	Mornings	27733		
4	Dawn	5242		

Insights :

An insightful observation from the query reveals distinct patterns in order placements based on different times of the day. Notably, the highest number of orders occur during the afternoon hours, between 13:00 and 18:00, totaling 38,135 orders. Following closely behind are orders placed during the night and morning periods, with counts of 28,331 and 27,733 respectively. Conversely, the dawn hours see the least order activity, with only 5,242 orders recorded.

Recommendation :

Based on these insights, we can optimize the revenue by strategically

- **Afternoon Focus:** Since the afternoon hours witness the highest order activity, the organization can prioritize promotional campaigns, product launches, or special offers during this period. Ensuring sufficient staffing and inventory levels during peak hours can help capitalize on increased demand and enhance customer satisfaction.
- **Night and Morning Engagement:** While order counts are slightly lower during night and morning periods, they still represent significant revenue opportunities. Implementing targeted marketing initiatives such as email campaigns, exclusive early morning or late-night discounts, or personalized promotions can encourage more sales during these times.
- **Dawn Optimization:** Despite experiencing the least order activity during dawn hours, there's potential to capture additional revenue by catering to early risers or night owls. Offering incentives like limited-time deals, boom discounts can attract customers during this period and increase sales.

Q3.Evolution of E-commerce orders in the Brazil region:

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

Query :

```
with month_seasonality as
(
  SELECT
    c.customer_state,
    Extract(MONTH from o.order_purchase_timestamp) as Month,
    FORMAT_DATETIME('%B',o.order_purchase_timestamp) as Month_name,
    COUNT(order_id) as
order_count_over_month_on_month_in_each_state,
  FROM
    `target.customers` c
  JOIN
    `target.orders` o USING(customer_id)
  GROUP BY
    c.customer_state,
    Extract(MONTH from order_purchase_timestamp),
    FORMAT_DATETIME('%B',o.order_purchase_timestamp)
)
SELECT
  customer_state,
  Month,
  Month_name,
  order_count_over_month_on_month_in_each_state
FROM
  Month_seasonality
ORDER BY
  Month,
  order_count_over_month_on_month_in_each_state desc
LIMIT
  10;
```

Output :

← Query results					
JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
		EXECUTION GRAP			
Row	customer_state ▼	Month ▼	Month_name ▼	order_count_over_my	
1	SP	1	January	3351	
2	RJ	1	January	990	
3	MG	1	January	971	
4	PR	1	January	443	
5	RS	1	January	427	
6	SC	1	January	345	
7	BA	1	January	264	
8	GO	1	January	164	
9	ES	1	January	159	
10	DF	1	January	151	

Insights :

Here, from the above query we can get the month on month order placements in each state, where in the month of Jan, the state SP leads with more order placements of 3351 order count and followed by other states with its month on month order counts.

Recommendation :

Focus on High-Performing States: Given that the state SP leads in order placements for the month of January, the organization should allocate additional resources and marketing efforts towards this region. By leveraging the existing momentum and customer engagement, the organization can further strengthen its market presence and capitalize on the high demand.

Tailored Marketing Strategies: Utilize the insights on month-on-month order placements to develop tailored marketing campaigns for each state. By understanding the varying preferences and behaviors of customers in different regions, 'Target' can create targeted promotions, discounts, or product offerings that resonate with local audiences, thereby driving increased sales.

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

Query :

```
SELECT
    customer_state,
    COUNT(customer_id) as customers_count_across_states
FROM
    `target.customers`
GROUP BY
    Customer_state
ORDER BY
    customers_count_across_states desc,
    Customer_state
LIMIT
    10;
```

Output :

Query results			
JOB INFORMATION		RESULTS	CHART
JSON			
Row	customer_state	customers_count_ac	
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	

Insights :

An insightful observation from the data reveals that the state 'SP' leads with the highest number of customers, with a significant customer count of 41,746. Following 'SP', the states 'RJ', 'MG', 'RS', and others also exhibit substantial customer counts. This insight underscores the importance of prioritizing marketing efforts and tailoring strategies to cater to the large customer base in 'SP', while also recognizing the potential opportunities presented by customers in other states with notable customer counts.

Recommendation :

Targeted Marketing in SP: Allocate a significant portion of marketing resources towards the state 'SP' to capitalize on the large customer base. Tailor marketing campaigns, promotions, and advertisements to resonate with the preferences and behaviors of customers in this region. By effectively engaging with SP customers, the organization can drive increased sales and revenue.

Expand Reach in Other States: While prioritizing 'SP', it's crucial for the organization to also recognize the potential opportunities presented by customers in other states with substantial customer counts such as 'RJ', 'MG', and 'RS'. Develop targeted marketing strategies to expand reach and capture market share in these regions. This may involve localized campaigns, partnerships with local businesses, or customization of products/services to meet regional preferences.

Customer Retention Strategies: Implement customer retention initiatives to foster loyalty and repeat purchases among existing customers in 'SP' and other high-density states. Offer personalized discounts, loyalty programs, or exclusive perks to incentivize repeat business and enhance customer lifetime value.

Q4. 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 avg_cte1 as
(
SELECT
    rank() over(order by AVG(p1.payment_value)) as rank_,
    AVG(p1.payment_value) as avg_cost_of_orders_2017
FROM
    `target.payments` p1
    JOIN
    `target.orders` o1 USING(order_id)
WHERE
    o1.order_purchase_timestamp BETWEEN '2017-01-01 00:00:00' AND
'2017-08-31 23:59:59'
),
avg_cte2 as
(
SELECT
    rank() over(order by AVG(p2.payment_value)) as rank_,
    AVG(p2.payment_value) as avg_cost_of_orders_2018
FROM
    `target.payments` p2
    JOIN
    `target.orders` o2 USING(order_id)
WHERE
    o2.order_purchase_timestamp BETWEEN '2018-01-01 00:00:00' AND
'2018-08-31 23:59:59'
)
SELECT
    ROUND(((ac2.avg_cost_of_orders_2018 -
ac1.avg_cost_of_orders_2017) / ac1.avg_cost_of_orders_2017) * 100,4) as
Percentage_increase_cost_of_orders_2017_2018
```

```
FROM
    avg_cte1 ac1
    JOIN
    avg_cte2 ac2 USING(rank_);
```

Output :

Query results		
JOB INFORMATION		RESULT
Row	Percentage_increase	
1	3.2253	

Insights :

An insightful observation gleaned from our data analysis reveals a notable increase in the cost of orders from the year 2017 to 2018, specifically between the months of January and August. This period reflects a significant percentage increase of 3.225%, indicating a substantial rise in order costs over the specified timeframe.

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

Query :

```
SELECT
    c.customer_state,
    ROUND(SUM(payment_value),4) as Total_price_of_order_placed,
    ROUND(AVG(payment_value),4) as Average_price_of_order_placed
FROM
    `target.customers` c
    JOIN
    `target.orders` USING(customer_id)
    JOIN
    `target.payments` USING(order_id)
GROUP BY
    c.customer_state
LIMIT 10;
```

Output :

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION D
Row	customer_state ▼	Total_price_of_order	Average_price_of_order		
1	RJ	2144379.69	158.5259		
2	RS	890898.54	157.1804		
3	SP	5998226.96	137.5046		
4	DF	355141.08	161.1348		
5	PR	811156.38	154.1536		
6	MT	187029.29	195.2289		
7	MA	152523.02	198.8566		
8	AL	96962.06	227.0774		
9	MG	1872257.26	154.7064		
10	PE	324850.44	187.9922		

Insights :

An insightful revelation from the data highlights the disparity in revenue generation among states, as determined by the total and average prices of orders. By examining these metrics, we can discern which states yield substantial profits and which lag behind in revenue generation. This insight underscores the importance of analyzing state-level revenue performance to inform strategic decisions aimed at maximizing profitability and addressing areas of opportunity within the organization.

Recommendation :

Focus on High-Profit States: Allocate resources and marketing efforts towards states that yield substantial profits based on total and average prices of orders. Develop targeted marketing campaigns and promotional strategies to capitalize on existing customer base and enhance sales in these regions.

Identify Growth Opportunities: Analyze states with lower revenue generation to identify growth opportunities and potential untapped markets. Explore factors contributing to lower sales, such as market saturation, competition, or lack of brand visibility.

Optimize Pricing and Product Offerings: Review pricing strategies and product assortments to align with state-level revenue performance. Adjust pricing tiers or introduce new product variants tailored to the preferences and purchasing power of customers in different states. Continuously monitor pricing elasticity and customer feedback to optimize revenue potential.

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

Query :

```
SELECT
    customer_state,
    ROUND(SUM(freight_value),4) as Total_freight_value_of_each_state,
    ROUND(AVG(freight_value),4) as
                                   Average_freight_value_of_each_state
FROM
    `target.customers`
JOIN
    `target.orders` USING(customer_id)
JOIN
    `target.order_items` USING(order_id)
GROUP BY
    Customer_state
ORDER BY
    Customer_state
LIMIT
    10;
```

Output :

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETA
Row	customer_state ▼	Total_freight_value_c	Average_freight_valu		
1	AC	3686.75	40.0734		
2	AL	15914.59	35.8437		
3	AM	5478.89	33.2054		
4	AP	2788.5	34.0061		
5	BA	100156.68	26.364		
6	CE	48351.59	32.7142		
7	DF	50625.5	21.0414		
8	ES	49764.6	22.0588		
9	GO	53114.98	22.7668		
10	MA	31523.77	38.257		

Insights :

The data reveals the total and average freight values for shipments originating from each state. By identifying the states with the highest and lowest total and average freight values, the organization can optimize logistics and shipping strategies. This insight has the potential to drive cost savings by facilitating more

efficient routing, carrier selection, and inventory management practices, ultimately contributing to improved operational efficiency and profitability.

Recommendation :

Optimize Shipping Routes: Identify states with lower total and average freight values as potential hubs for centralized distribution centers. By strategically locating distribution centers closer to regions with higher shipping costs, the organization can minimize transportation expenses and improve delivery efficiency, ultimately reducing shipping overhead and enhancing revenue margins.

Negotiate Carrier Contracts: Leverage insights on freight costs to negotiate favorable contracts with shipping carriers. Engage in discussions with carriers to secure discounted rates or volume-based pricing agreements, especially for states with higher shipping expenses. Lowering shipping costs through negotiated contracts can increase profit margins and competitiveness in the market.

Implement Dynamic Pricing: Adjust product pricing based on shipping costs to maximize profitability. Offer flexible pricing strategies, such as zone-based or weight-based shipping rates, to align with varying freight expenses across different states. By accurately reflecting shipping costs in product prices, the organization can optimize revenue while remaining competitive in the market.

Q5. 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:

■ $\text{time_to_deliver} = \text{order_delivered_customer_date} - \text{order_purchase_timestamp}$

■ $\text{diff_estimated_delivery} = \text{order_delivered_customer_date} - \text{order_estimated_delivery_date}$

Query :

```
SELECT
    Order_id,
    DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date,
              DAY) as Difference_between_estimed_and_actual_delivery,
    DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY)
              as No_of_day_difference_between_delivery_and_purchase_date
FROM
    `target.orders`;
```

Output :

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETA
Row	order_id	Difference_between	No_of_day_differenc		
1	1950d777989f6a877539f5379...	12	30		
2	2c45c33d2f9cb8ff8b1c86cc28...	-28	30		
3	65d1e226dfaeb8cdc42f66542...	-16	35		
4	635c894d068ac37e6e03dc54e...	-1	30		
5	3b97562c3aee8bdecb5c2e45...	0	32		
6	68f47f50f04c4cb6774570cfde...	-1	29		
7	276e9ec344d3bf029ff83a161c...	4	43		
8	54e1a3c2b97fb0809da548a59...	4	40		
9	fd04fa4105ee8045f6a0139ca5...	1	37		
10	302bb8109d097a9fc6e9cefc5...	5	33		

Insights :

An insightful observation derived from this query highlights the time taken in days to deliver products from the purchase date, as well as the time taken between the estimated and actual delivery dates. This insight provides valuable metrics for evaluating the organization's fulfillment efficiency and accuracy in meeting

delivery commitments. Analyzing these timeframes can uncover potential areas for improvement in logistics, inventory management, and order processing to enhance customer satisfaction and optimize operational performance.

Recommendation :

Streamline Fulfillment Processes: Evaluate and streamline fulfillment processes to minimize the time taken to deliver products from the purchase date. Identify inefficiencies in order processing, packaging, and shipping, and implement measures to expedite these processes. Faster order fulfillment can lead to improved customer satisfaction, increased repeat purchases, and ultimately higher revenue.

Inventory Management: Optimize inventory management practices to ensure timely availability of products for order fulfillment. Implement demand forecasting techniques to accurately predict customer demand and maintain adequate stock levels. By reducing stockouts and backorders, the organization can minimize delays in delivery and enhance customer satisfaction, leading to higher revenue.

Improve Shipping Efficiency: Analyze the variance between estimated and actual delivery dates to identify opportunities for improving shipping efficiency. Evaluate carrier performance, transit times, and delivery routes to minimize delays and optimize delivery schedules. Negotiate favorable contracts with shipping carriers to ensure reliable and cost-effective delivery services, enhancing customer satisfaction and loyalty

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

Query :

```
with cte as
(
SELECT
    c.customer_state,
    AVG(freight_value) as avg_freight_value
FROM
    `target.customers` c
    JOIN
    `target.orders` USING(customer_id)
```

```

        JOIN
        `target.order_items` USING(order_id)
    GROUP BY c.customer_state
),
cte2 as
(
    SELECT
        customer_state,
        avg_freight_value as top_5_avg_freight_value,
        rank() over(order by avg_freight_value desc) as rank_
    FROM
        cte
    ORDER BY
        rank_
    LIMIT
        5
),
cte3 as
(
    SELECT
        customer_state,
        avg_freight_value as least_5_avg_freight_value,
        rank() over(order by avg_freight_value ) as rank_
    FROM
        cte
    ORDER BY
        rank_
    LIMIT
        5
)
SELECT
    c2.customer_state as top_5_states_avg_freight_value,
    c2.top_5_avg_freight_value,
    c3.customer_State as least_5_states_avg_freight_value,
    c3.least_5_avg_freight_value
FROM

```

```
cte2 c2
JOIN
cte3 c3 USING(rank_)
;
```

Output :

Query results				
JOB INFORMATION		RESULTS	CHART	JSON
		EXECUTION DETAILS		
		EXECUTION GRAI		
Row	top_5_states_avg_freight_value	top_5_avg_freight_vg	least_5_states_avg_freight_value	least_5_avg_freight_v
1	RR	42.98442307692...	SP	15.14727539041...
2	PB	42.72380398671...	PR	20.53165156794...
3	RO	41.06971223021...	MG	20.63016680630...
4	AC	40.07336956521...	RJ	20.96092393168...
5	PI	39.14797047970...	DF	21.04135494596...

Insights :

An insightful observation from the data highlights the states with their respective freight values, showcasing the top and least five states based on average freight value. This insight provides valuable information on the distribution of shipping costs across different states, allowing the organization to identify regions with the highest and lowest average freight expenses. Analyzing these trends can inform strategic decisions related to pricing, inventory management, and logistics optimization, ultimately contributing to more efficient operations and potentially higher profit margins.

Recommendation :

Strategic Pricing Adjustments: Tailor pricing strategies based on the average freight values in different states. Consider adjusting product pricing or offering regional promotions to offset higher shipping costs in states with elevated freight expenses. By aligning prices with shipping costs, the organization can maintain competitiveness while maximizing profitability.

Optimize Inventory Distribution: Strategically distribute inventory to minimize shipping distances and costs. Prioritize stocking popular products in regions with lower freight expenses to reduce overall transportation costs. Implement demand forecasting and inventory replenishment strategies to ensure adequate stock levels while minimizing inventory holding costs.

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

Query :

```
with date_diff_cte as
(
    SELECT
        order_id,
        customer_id,

        DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp,
        day) as delivery_time
    FROM
        `target.orders`
),
avg_del_time as
(
    SELECT
        customer_state,
        AVG(date_diff_cte.delivery_time) as avg_delivery_time
    FROM
        date_diff_cte
    JOIN
        `target.customers` USING(customer_id)
    GROUP BY
        customer_state
),
top_5_states as
(
    SELECT
        customer_state as states_with_highest_delivery_time ,
        avg_delivery_time as top_5_avg_delivery_time,
        rank() over(order by avg_delivery_time desc) as rank_
    FROM
        avg_del_time
    ORDER BY
        rank_
    LIMIT
```

```

5
),
least_5_states as
(
SELECT
    customer_state as states_with_lowest_delivery_time,
    avg_delivery_time as least_5_avg_delivery_time,
    rank() over(order by avg_delivery_time) as rank_
FROM
    avg_del_time
ORDER BY
    rank_
LIMIT
5
)
SELECT
    t5.states_with_highest_delivery_time,
    CEIL(t5.top_5_avg_delivery_time) as highest_5_avg_delivery_time,
    l5.states_with_lowest_delivery_time,
    CEIL(l5.least_5_avg_delivery_time) as lowest_5_avg_delivery_time
FROM
    top_5_states t5
    JOIN
    least_5_states l5 USING(rank_)
;

```

Output :

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GR.
Row	states_with_highest_delivery_time	highest_5_avg_delive	states_with_lowest_delivery_time	lowest_5_avg_deliver		
1	RR	29.0	SP	9.0		
2	AP	27.0	PR	12.0		
3	AM	26.0	MG	12.0		
4	AL	25.0	DF	13.0		
5	PA	24.0	SC	15.0		

Insights :

An insightful observation from the data showcases the top ten states with the shortest average delivery times, led by state 'SP' with an average delivery time of 9 days, and the bottom ten states with the longest average delivery times, with state 'RR' recording the highest delivery time of 29 days. This insight underscores the significant variability in delivery efficiency across different states.

Recommendation :

Optimize Delivery Routes and Logistics: Analyze the factors contributing to longer delivery times in states with higher average delivery times, such as inefficient routing, inadequate transportation infrastructure, or limited fulfillment capabilities. Optimize delivery routes, explore alternative transportation options, and invest in logistics technology to streamline operations and reduce transit times. By improving delivery efficiency, the organization can enhance customer satisfaction, increase repeat purchases, and drive revenue growth.

Prioritize High-Performing States: Allocate resources and prioritize initiatives in states with shorter average delivery times, such as 'SP', to maintain high service levels and capitalize on operational efficiencies. Implement targeted marketing campaigns to encourage repeat purchases in these regions. By nurturing relationships with customers in high-performing states, the organization can maximize revenue opportunities and enhance long-term profitability.

Invest in Customer Communication and Transparency: Enhance communication with customers regarding delivery expectations, updates, and potential delays. Provide real-time tracking information, delivery notifications and proactive updates to keep customers informed and manage expectations effectively. Transparency and proactive communication can mitigate dissatisfaction caused by longer delivery times, improve trust and positive customer experiences, leading to increased loyalty and revenue.

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 :

```
with date_diff_cte as
(
    SELECT
        order_id,
        customer_id,

        DATE_DIFF(order_delivered_customer_date,order_estimated_delivery_date, day) as delivery_time
    FROM
        `target.orders`
),
avg_del_time as
(
    SELECT
        customer_state,
        AVG(delivery_time) as avg_delivery_time
    FROM
        date_diff_cte
    JOIN
        `target.customers` USING(customer_id)
    GROUP BY
        customer_state
)
SELECT
    customer_state as top5_states_with_early_delivery_than_estimated
FROM
    avg_del_time
ORDER BY
    avg_delivery_time
LIMIT
    5
;
```

Output :

←

Query results

JOB INFORMATION		RESULTS	CI
Row	top5_states_with_early_delivery_th		
1	AC		
2	RO		
3	AP		
4	AM		
5	RR		

Insights :

An insightful observation from the data highlights the states where deliveries consistently arrive earlier than the estimated delivery time provided to the customer. This insight reflects operational efficiency and effectiveness in meeting or exceeding customer expectations for delivery timelines.

Recommendation :

Leverage Early Delivery as a Competitive Advantage: Highlight early delivery performance as a key differentiator in marketing efforts and customer communications. Emphasize the organization's commitment to exceeding customer expectations and delivering exceptional service by consistently providing early deliveries.

Promote Premium Shipping Options: Introduce premium shipping options that guarantee early or expedited delivery for customers willing to pay a premium fee. Position these premium shipping services as a value-added proposition, offering convenience and peace of mind to customers who prioritize speed and reliability

Collect and Analyze Customer Feedback: Solicit feedback from customers regarding their experience with early deliveries and use this information to further enhance service levels. Gather insights into customer preferences, expectations, and areas for improvement related to delivery speed and reliability.

Q6. Analysis based on the payments:

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

Query:

```
SELECT
    EXTRACT(MONTH from order_purchase_timestamp) as month_number,
    FORMAT_TIMESTAMP('%B', DATE(order_purchase_timestamp)) as
month_name,
    payment_type,
    COUNT(order_id) as order_count
FROM
    `target.orders`
    JOIN
    `target.payments` USING(order_id)
GROUP BY
    month_number,
    month_name,
    payment_type
ORDER BY
    month_number,
    payment_type
;
```

Output:

Query results					
JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
				EXECUTION GF	
Row	month_number	month_name	payment_type	order_count	
1	1	January	UPI	1715	
2	1	January	credit_card	6103	
3	1	January	debit_card	118	
4	1	January	voucher	477	
5	2	February	UPI	1723	
6	2	February	credit_card	6609	
7	2	February	debit_card	82	
8	2	February	voucher	424	
9	3	March	UPI	1942	
10	3	March	credit_card	7707	

Insights:

This data reveals the month-on-month order payment types preferred by customers. By comparing the data, it becomes evident that the majority of customers opt to make their orders using credit cards, followed by UPI (Unified Payments Interface), vouchers, and debit cards. This insight underscores the importance of offering diverse payment options to accommodate customer preferences and enhance convenience, ultimately contributing to improved customer satisfaction and loyalty.

Recommendations:

Offer Discounts: Develop targeted promotional campaigns that incentivize customers to use the most and least used payment types. For the most used payment type, offer exclusive discounts, cashback rewards, or bonus points to encourage continued usage and loyalty. Conversely, for the least used payment type, provide special offers, limited-time promotions, or introductory discounts to incentivize adoption and increase usage rates.

Optimize Payment Processing: Streamline and optimize payment processing systems to accommodate the most popular payment methods such as credit cards and UPI. Ensure seamless integration with payment gateways and platforms to facilitate smooth and secure transactions. By reducing friction during the checkout process, the organization can minimize abandoned carts and increase conversion rates, ultimately driving revenue growth.

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

Query:

```
SELECT
    payment_installments,
    COUNT(order_id) order_count
FROM
    `target.payments`
GROUP BY
    payment_installments
```

Output:

Query results			
JOB INFORMATION		RESULTS	CHART
Row	payment_installment	order_count	
1	0	2	
2	1	52546	
3	2	12413	
4	3	10461	
5	4	7098	
6	5	5239	
7	6	3920	
8	7	1626	
9	8	4268	
10	9	644	

Insights:

An insightful observation derived from the data reveals the number of orders made through installments on a monthly basis. This insight sheds light on the payment preferences and behavior of customers, indicating their willingness to spread payments over time rather than making a single upfront payment.

Recommendations:

Promote Installment Payment Options: Capitalize on customers' preference for spreading payments over time by actively promoting installment payment options. Highlight the flexibility and convenience of installment plans in marketing communications, product listings, and checkout processes. By emphasizing the affordability and accessibility of installment payments, the organization can attract more customers and increase sales, ultimately driving revenue growth.