

## **Business Case Study:**

Target SQL

### **Submitted by:**

Jay Abotiawala

### **Problem Statement:**

- Assuming you are a data analyst/ scientist at Target, you have been assigned the task of analysing the given dataset to extract valuable insights and provide actionable recommendations.

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_casestudy.INFORMATION_SCHEMA.COLUMNS
WHERE
    table_name = 'customers'
```

- Output:

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

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

- Query:

```
SELECT
    MIN(order_purchase_timestamp) AS min_orderplacement_time,
    MAX(order_purchase_timestamp) AS max_orderplacement_time
FROM `target_casestudy.orders`
```

- Output:

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

- Actionable Insights & Recommendations:

- ❖ The dataset comprises a range of orders recorded from September 2016 to October 2018.

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

- Query:

- Case1:

```
SELECT
  COUNT(DISTINCT customer_city) AS customer_city_count,
  COUNT(DISTINCT customer_state) AS customer_state_count
FROM `target_casestudy.customers`
```

- Case2:

```
SELECT
  COUNT(DISTINCT seller_city) AS seller_city_count,
  COUNT(DISTINCT seller_state) AS seller_state_count
FROM `target_casestudy.sellers`
```

- Output:

- Case1:

Row	customer_city_count	customer_state_count
1	4119	27

- Case2:

Row	seller_city_count	seller_state_count
1	611	23

- Actionable Insights & Recommendations:

- ❖ Customers are spread across a wide range of cities and states, indicating a diverse customer population. On the other hand, sellers are relatively concentrated in a smaller number of cities, but still cover a diverse range of states.
- ❖ To improve the presence of sellers in states with a gap in customer coverage, the company should:
  - Identify potential sellers in the gap states
  - Partner with local businesses to expand reach
  - Expand marketing efforts in the gap states
  - Gather and act on customer feedback

## 2. 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 order_year,
  count(*) as number_of_orders
FROM `target_casestudy.orders`
group by order_year
order by order_year
```

- Output:

Row	order_year	number_of_orders
1	2016	329
2	2017	45101
3	2018	54011

- Actionable Insights & Recommendations:

- ❖ Based on data, there is a clear trend of increasing orders placed over the past years. The number of orders more than doubled (considering  $329 \times 3 = 987$  projected orders for whole year-2016) from 2016 to 2017, and there was further growth in 2018.
- ❖ This indicates a positive trend of growing customer demand or business expansion over time.

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

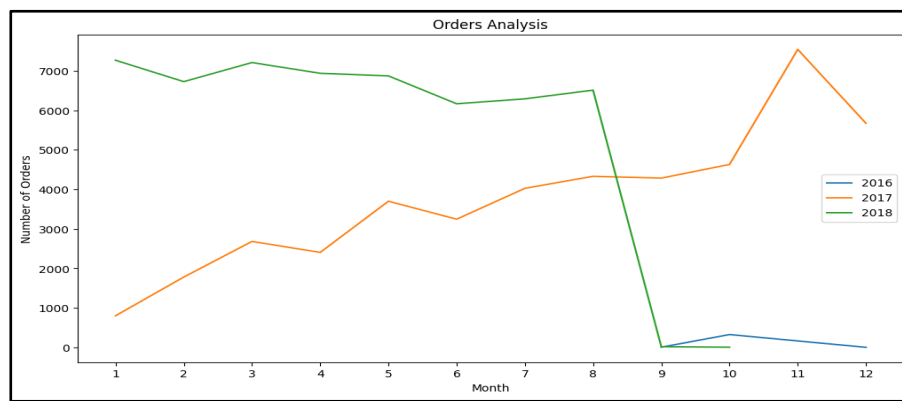
- Query:

```
SELECT
  EXTRACT(MONTH FROM order_purchase_timestamp) AS order_month,
  EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
  COUNT(*) AS number_of_orders
FROM `target_casestudy.orders`
GROUP BY order_year, order_month
ORDER BY order_year, order_month
```

- **Output:**

Row	order_month	order_year	number_of_orders
1	9	2016	4
2	10	2016	324
3	12	2016	1
4	1	2017	800
5	2	2017	1780
6	3	2017	2682
7	4	2017	2404
8	5	2017	3700
9	6	2017	3245
10	7	2017	4026

- **Actionable Insights & Recommendations:**



- ❖ To address the monthly seasonality in order placements, businesses can consider the following strategies based on the observed decrease in orders during April, June, and December:
  - Plan for seasonal promotions and exclusive discounts during holiday season.
  - Introduce limited-time products to create a sense of urgency.
  - Utilize social media and email marketing to promote holiday offers.

**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'
    WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 19
AND 23 THEN 'Night'
  END AS time_period,
  COUNT(*) AS num_of_orders,
FROM `target_casestudy.orders`
GROUP BY time_period
ORDER BY num_of_orders DESC
```

- **Output:**

Row	time_period	num_of_orders
1	Afternoon	38135
2	Night	28331
3	Morning	27733
4	Dawn	5242

- **Actionable Insights & Recommendations:**

- ❖ 'Afternoon' is the most popular time period.
- ❖ 'Night' is the second busiest time period followed by 'Morning' time period.
- ❖ 'Dawn' has relatively lower order counts.
- ❖ It can be valuable for business in terms of scheduling promotions, optimizing customer service, and managing resources to align with the peak order placement hours.
- ❖ To increase orders during the dawn time period, the company should:
  - Offer exclusive discounts or promotions during dawn
  - Personalize recommendations based on customer preferences during dawn
  - Implement a loyalty program with rewards for dawn orders

### 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
    c.customer_state,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
    COUNT(o.order_id) AS no_of_orders,
FROM
    `target_casestudy.orders` o
    JOIN `target_casestudy.customers` c ON o.customer_id =
c.customer_id
GROUP BY
    month, customer_state
ORDER BY
    customer_state, month
```

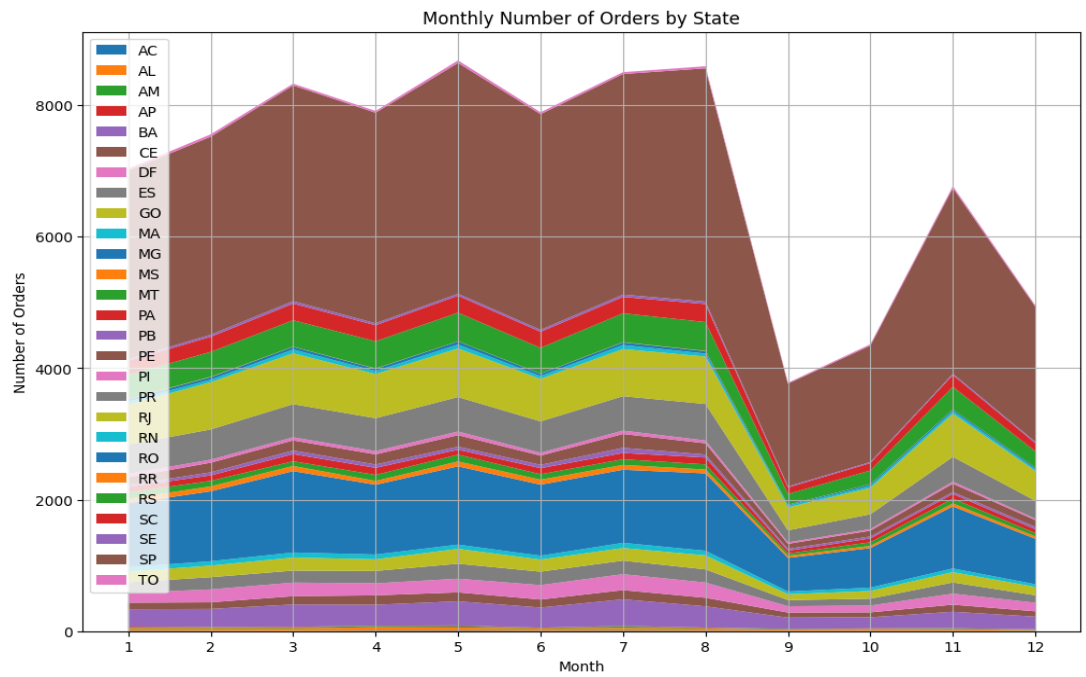
- Output:

Row	customer_state	month	no_of_orders
1	AC	1	8
2	AC	2	6
3	AC	3	4
4	AC	4	9
5	AC	5	10
6	AC	6	7
7	AC	7	9
8	AC	8	7
9	AC	9	5
10	AC	10	6

- Actionable Insights & Recommendations:

- ❖ There is a consistent increase in orders from January to August, with a peak in August, and then a decline in the following month.
- ❖ The number of orders varies significantly across different states. The state of "SP" consistently has the highest number of orders each month, followed by "MG," "RJ," and "PR." These states consistently have a higher volume of orders compared to other states.
- ❖ On the other hand, states like "RR," "TO," "AM," and "AP" consistently have lower order volumes throughout the year.
- ❖ States from the southeastern region of Brazil (e.g., "SP," "MG," "RJ") consistently have higher order volumes compared to states from the

northern (e.g., "AM," "AP") and northeastern (e.g., "MA," "PB") regions.



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

- Query:**

```
SELECT
    c.customer_state,
    COUNT(distinct o.customer_id) AS no_of_customers
FROM
    `target_casestudy.orders` o
JOIN `target_casestudy.customers` c ON o.customer_id =
    c.customer_id
GROUP BY
    customer_state
ORDER BY
    no_of_customers DESC
```

- Output:**

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



- **Actionable Insights & Recommendations:**

- ❖ The state with the highest number of customers is São Paulo (SP) with 41,746 customers. This indicates a significant customer base in São Paulo, which is the most populous state in Brazil and has a strong economy.
- ❖ Following São Paulo, the states with substantial customer bases are Rio de Janeiro (RJ) with 12,852 customers and Minas Gerais (MG) with 11,635 customers. These states also have large populations and major urban centres, contributing to their higher customer numbers.
- ❖ Overall, the distribution of customers across the states reflects population sizes, economic activities, and regional characteristics.

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:

```
SELECT
    ROUND(SUM(CASE WHEN order_year = 2017 THEN cost_of_orders END),2)
AS cost_of_orders_2017,
    ROUND(SUM(CASE WHEN order_year = 2018 THEN cost_of_orders END),2)
AS cost_of_orders_2018,
    ROUND(((SUM(CASE WHEN order_year = 2018 THEN cost_of_orders END)-
    SUM(CASE WHEN order_year = 2017 THEN cost_of_orders END))/SUM(CASE
    WHEN order_year = 2017 THEN cost_of_orders END)*100),2) AS
    percentage_increase_in_cost
FROM
(
    SELECT
        extract(year from o.order_purchase_timestamp) as order_year,
        extract(month from o.order_purchase_timestamp) as order_month,
        sum(p.payment_value) as cost_of_orders
    FROM
        `target_casestudy.orders` o
    JOIN `target_casestudy.payments` p ON o.order_id = p.order_id
    WHERE
        EXTRACT(YEAR FROM o.order_purchase_timestamp) IN (2017, 2018)
        AND
        EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
    GROUP BY order_year, order_month
    ORDER BY order_year, order_month
) X
```

- Output:

Row	cost_of_orders_2017	cost_of_orders_2018	percentage_increase_in_cost
1	3669022.12	8694733.84	136.98

- Actionable Insights & Recommendations:

- ❖ The substantial increase in the total cost of orders from 2017 to 2018 suggests a positive growth trend. It indicates that the business experienced significant growth in terms of customer orders and revenue generated.

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

- **Query:**

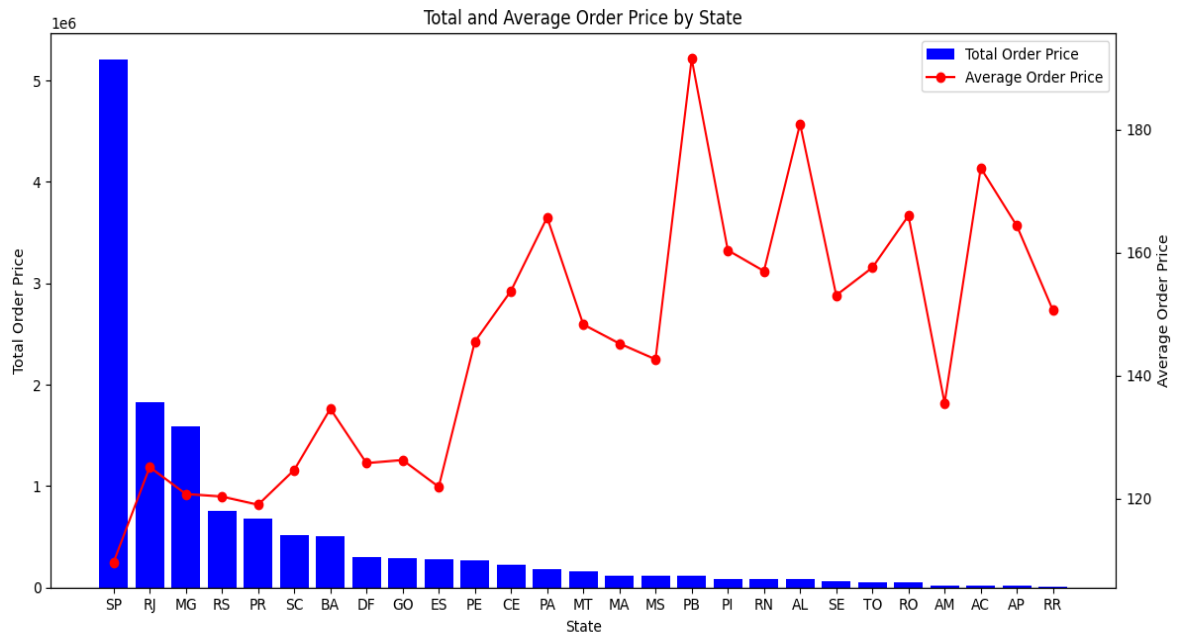
```
SELECT c.customer_state AS state,
       ROUND(SUM(oi.price),2) AS total_order_price,
       ROUND(AVG(oi.price),2) AS average_order_price
FROM `target_casestudy.customers` AS c
JOIN `target_casestudy.orders` AS o ON c.customer_id = o.customer_id
JOIN `target_casestudy.order_items` AS oi ON o.order_id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

- **Output:**

Row	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

- **Actionable Insights & Recommendations:**

- ❖ São Paulo has the highest overall order value among all the states.
- ❖ The state with the highest average order price is Paraíba (PB) with an average of 191.48 per order. This suggests that customers from Paraíba tend to spend more on their individual orders compared to other states.
- ❖ Targeting marketing campaigns or offering tailored promotions to customers in states with lower average order prices (e.g., Amazonas, Roraima, and Amapá) could help increase their spending and overall business revenue.



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

- Query:**

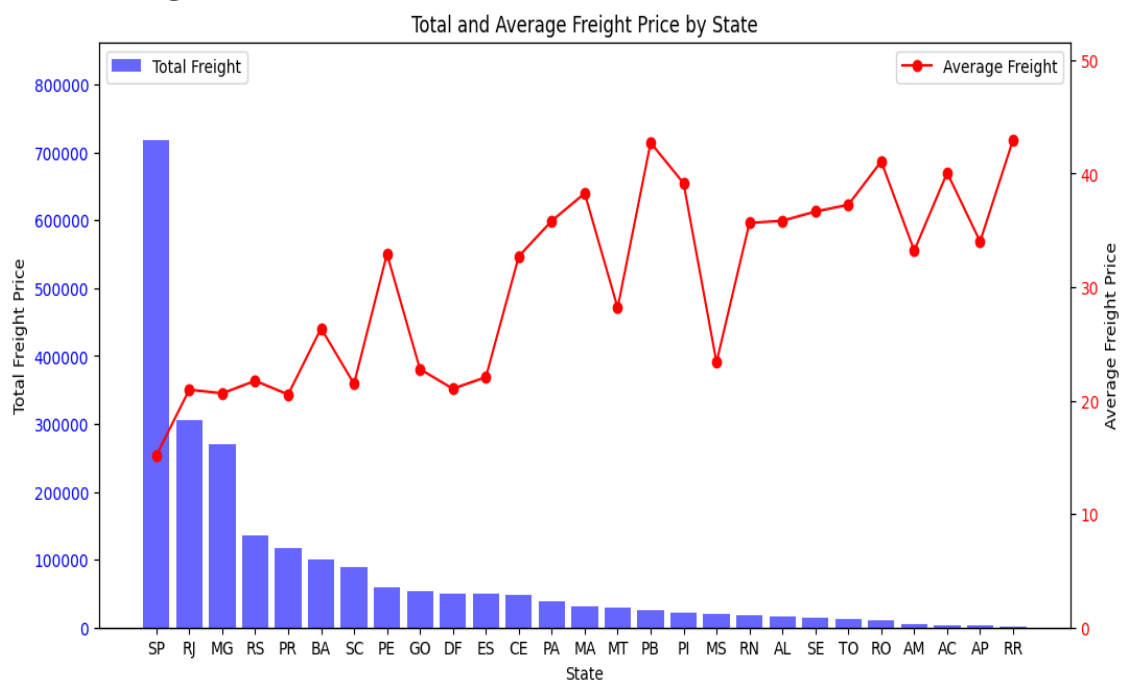
```
SELECT c.customer_state AS state,
       ROUND(SUM(oi.freight_value),2) AS total_freight_price,
       ROUND(AVG(oi.freight_value),2) AS average_freight_price
FROM `target_casestudy.customers` AS c
JOIN `target_casestudy.orders` AS o ON c.customer_id = o.customer_id
JOIN `target_casestudy.order_items` AS oi ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY total_freight_price DESC
```

- Output:**

Row	state	total_freight_price	average_freight_price
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

- **Actionable Insights & Recommendations:**

- ❖ São Paulo (SP) has the highest overall shipping volume compared to other states.
- ❖ There is a notable variation in average freight prices among states. These differences could be attributed to factors such as distance, transportation infrastructure, and logistics efficiency.
- ❖ States with higher average freight prices, such as Paraíba (PB), Piauí (PI), and Maranhão (MA), may have longer shipping distances or face logistical challenges, leading to higher shipping costs. On the other hand, states with lower average freight prices, like São Paulo (SP) and Rio de Janeiro (RJ), may benefit from better transportation infrastructure and economies of scale due to higher shipping volumes.
- ❖ Analyzing the shipping patterns and factors affecting freight prices in states with higher costs may uncover areas where efficiency improvements or negotiation with carriers could lead to cost savings.



## 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:**

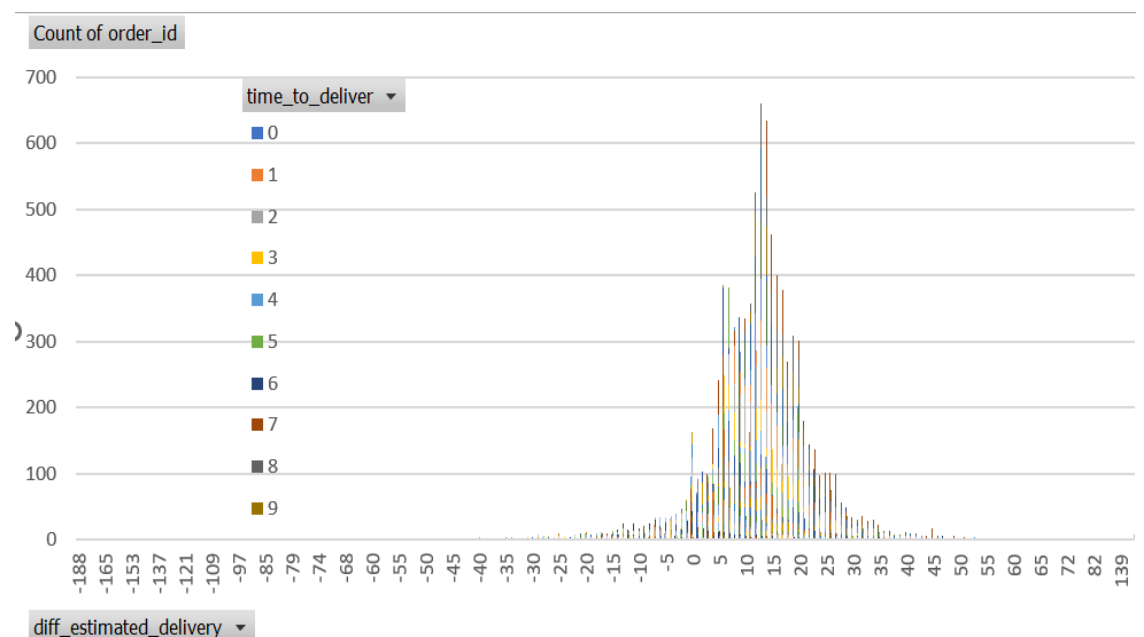
```
SELECT
    order_id,
    date_diff(order_delivered_customer_date, order_purchase_timestamp,
    day) AS time_to_deliver,
    date_diff(order_estimated_delivery_date,
    order_delivered_customer_date, day) AS diff_estimated_delivery
FROM `target_casestudy.orders`
WHERE order_status = 'delivered'
ORDER BY order_id
```

- **Output:**

Row	order_id	time_to_deliver	diff_estimated_delivery
1	00010242fe8c5a6d1ba2dd792...	7	8
2	00018f77f2f0320c557190d7a1...	16	2
3	000229ec398224ef6ca0657da...	7	13
4	00024acbcd0a6daa1e931b03...	6	5
5	00042b26cf59d7ce69dfabb4e...	25	15
6	00048cc3ae777c65dbb7d2a06...	6	14
7	00054e8431b9d7675808bcb8...	8	16
8	000576fe39319847cbb9d288c...	5	15
9	0005a1a1728c9d785b8e2b08b...	9	0
10	0005f50442cb953dcd1d21e1f...	2	18

- **Actionable Insights & Recommendations:**

- ❖ Based on the chart, it appears that the majority of orders are delivered within the estimated time. However, it is important for the business to address orders that could not be delivered on time.
- ❖ To ensure customer satisfaction, it is crucial to conduct a thorough root cause analysis in collaboration with the team. This analysis will help identify any constraints or issues that led to delays and enable the business to resolve them effectively.
- ❖ Proper Root Cause Analysis to be done along with team to understand constraints and resolve it to increase customer satisfaction.



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

- **Query:**

```
(SELECT c.customer_state AS state, ROUND(AVG(oi.freight_value), 2) AS
average_freight_price
FROM `target_casestudy.customers` AS c
JOIN `target_casestudy.orders` AS o ON c.customer_id = o.customer_id
JOIN `target_casestudy.order_items` AS oi ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY average_freight_price DESC
LIMIT 5
)
UNION ALL
(
SELECT c.customer_state AS state, ROUND(AVG(oi.freight_value), 2) AS
average_freight_price
```

```

FROM `target_casestudy.customers` AS c
JOIN `target_casestudy.orders` AS o ON c.customer_id = o.customer_id
JOIN `target_casestudy.order_items` AS oi ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY average_freight_price
LIMIT 5)

```

- **Output:**

Row	state	average_freight_price
1	RR	42.98
2	PB	42.72
3	RO	41.07
4	AC	40.07
5	PI	39.15
6	SP	15.15
7	PR	20.53
8	MG	20.63
9	RJ	20.96
10	DF	21.04

Top 5 Highest Average  
Freight value

Top 5 Lowest Average  
Freight value

- **Actionable Insights & Recommendations:**

- ❖ The average freight values vary significantly across different states. It is important to conduct a detailed regional analysis to understand the factors contributing to these differences.
- ❖ The states with the highest average freight values (RR, PB, RO, AC, and PI) present potential market opportunities. Business can focus on these states to offer specialized logistics services or target industries that heavily rely on freight transportation.
- ❖ For states with the lowest average freight values (SP, PR, MG, RJ, and DF), business can focus on optimizing costs to remain competitive.
- ❖ Embracing technology solutions such as route optimization algorithms, real-time tracking systems, and data analytics can improve operational efficiency, reduce costs, and enhance overall customer experience.

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

- **Query:**

```

(SELECT
  c.customer_state AS state,

```



```

ROUND(AVG(date_diff(order_delivered_customer_date,
order_purchase_timestamp, day)),2) AS avg_time_to_deliver,
FROM `target_casestudy.customers` AS c
JOIN `target_casestudy.orders` AS o ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY avg_time_to_deliver DESC
LIMIT 5)
UNION ALL
(SELECT
c.customer_state AS state,
ROUND(AVG(date_diff(order_delivered_customer_date,
order_purchase_timestamp, day)),2) AS avg_time_to_deliver,
FROM `target_casestudy.customers` AS c
JOIN `target_casestudy.orders` AS o ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY avg_time_to_deliver
LIMIT 5)

```

- **Output:**

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

Top 5 Highest Average  
Delivery time

Top 5 Lowest Average  
Delivery time

- **Actionable Insights & Recommendations:**

- ❖ States with higher average delivery times may require improvements in transportation and logistics infrastructure to expedite the delivery process. Also analyze the delivery routes in these states to identify any inefficiencies and optimize them for faster and more efficient delivery.
- ❖ Implement tracking systems and provide real-time updates to customers, allowing them to track their deliveries.
- ❖ States with higher average delivery times, indicating that their logistics processes are well-optimized. Analyze their strategies to identify best practices that can be applied to other regions. Encourage collaboration and knowledge sharing between the logistics teams in these states and other regions to disseminate best practices and optimize delivery processes.

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
    X.state,
    ROUND((X.time_to_deliver-X.diff_estimated_delivery),2) AS
    order_delivery
FROM
    (SELECT
        c.customer_state as state,
        AVG(DATE_DIFF(o.order_delivered_customer_date,
o.order_purchase_timestamp, day)) AS time_to_deliver,
        AVG(DATE_DIFF(o.order_estimated_delivery_date,
o.order_delivered_customer_date, day)) AS diff_estimated_delivery
    FROM `target_casestudy.orders` o
    JOIN `target_casestudy.customers` c
    ON o.customer_id = c.customer_id
    GROUP BY c.customer_state) X
ORDER BY (X.time_to_deliver-X.diff_estimated_delivery)
LIMIT 5
```

- **Output:**

Row	state	order_delivery
1	SP	-1.84
2	PR	-0.84
3	MG	-0.75
4	RO	-0.22
5	AC	0.87

- **Actionable Insights & Recommendations:**

- ❖ By identifying the top-performing states for fast delivery and addressing any shortcomings in regions where delivery is slower, the company can enhance customer satisfaction, increase brand loyalty, and gain a competitive edge in the market.

## 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_casestudy.orders` o
JOIN `target_casestudy.payments` p ON o.order_id = p.order_id
WHERE p.payment_type != 'not_defined'
GROUP BY month, p.payment_type
ORDER BY month, order_count DESC
```

- Output:

Row	month	payment_type	order_count
1	1	credit_card	6103
2	1	UPI	1715
3	1	voucher	477
4	1	debit_card	118
5	2	credit_card	6609
6	2	UPI	1723
7	2	voucher	424
8	2	debit_card	82
9	3	credit_card	7707
10	3	UPI	1942

- Actionable Insights & Recommendations:

- ❖ The highest number of orders were placed using credit cards, followed by UPI, vouchers, and debit cards.
- ❖ The spike in order counts in July indicates a potential seasonal trend or a marketing campaign's success. The business can analyse the marketing activities during that period to understand the contributing factors and replicate them in other months.
- ❖ The gradual decrease in order counts from July to December suggests a need for targeted promotions or new product launches to maintain customer interest and engagement during these months.
- ❖ Debit card usage is relatively low. The business can investigate the reasons behind this and evaluate the feasibility of offering discounts or promotions for debit card users.

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

- Query:

```
SELECT payment_installments, COUNT(DISTINCT order_id) AS no_of_orders
FROM `target_casestudy.payments`
WHERE payment_installments > 0
GROUP BY payment_installments
```

- Output:

Row	payment_installments	no_of_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

- Actionable Insights & Recommendations:

- ❖ The majority of orders have fewer payment installments, with a significant number of orders falling into the range of 1 to 5 installments. This indicates that customers generally prefer to make payments over a shorter period of time.
- ❖ The data shows a decrease in the number of orders as the number of payment installments increases beyond a certain point (e.g., beyond 10 installments). Possible reasons could include increased interest rates or fees associated with higher installments, lack of customer awareness about extended payment options, or other barriers. Addressing these factors could help mitigate the drop-off and potentially increase sales.
- ❖ Analyzing customer feedback and default rates can help identify areas for improvement and fine-tune installment options to enhance customer satisfaction.

## **7. Actionable Insights & Recommendations:**

- ❖ Covered in individual question.