Business Case: Target SQL

Target Brazil E-Commerce Operational Performance Analysis (2016–2018)

Prepared by: Guntuku Hari Prasad

Introduction

Target is a globally recognized retail brand known for delivering value, innovation, and an exceptional customer experience. This case study focuses on analyzing Target's e-commerce operations in Brazil using transactional data from approximately 100,000 orders placed between 2016 and 2018.

The dataset provides a comprehensive view of the business across multiple operational dimensions including customer demographics, order patterns, pricing, freight costs, delivery performance, payment behavior, product characteristics, and customer feedback. Through structured SQL-based data analysis, this case study aims to uncover meaningful insights into the growth, efficiency, and economic impact of Target's operations in Brazil.

Objective of the Study

The primary objective of this analysis is to:
1. Understand order trends and seasonality patterns.
2. Analyze geographic distribution of customers and state-level performance.
3. Evaluate revenue growth and economic contribution.
4. Assess logistics efficiency through freight and delivery time analysis.
5. Examine payment patterns and installment behavior.
6. Generate actionable business recommendations for operational and strategic improvement.

Scope of Analysis

The study covers:
1. Exploratory Data Analysis (EDA)
2. Growth and seasonality trends
3. State-wise performance benchmarking
4. Revenue and freight economics
5. Delivery accuracy and logistics efficiency
6. Payment behavior analytics
7. Operational optimization insights

Analytical Approach

The analysis was conducted using SQL queries on structured relational datasets comprising customers, orders, order items, sellers, payments, reviews, products, and geolocation data. Metrics were carefully derived using aggregations, time-series analysis, and state-level segmentation to ensure reliable and actionable results.

Expected Outcome

By performing this analysis, we aim to:
1. Identify growth patterns in Brazil's e-commerce operations.
2. Detect regional performance strengths and bottlenecks.
3. Evaluate delivery efficiency versus customer expectations.
4. Understand customer payment preferences and installment usage.
5. Provide data-driven business recommendations to enhance profitability, operational scalability, and customer satisfaction.

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

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

```
SELECT
table_name,
column_name,
data_type
FROM `target-sql-487006.targetsql_01.INFORMATION_SCHEMA.COLUMNS`
where table_name = "customers";
```

Row	table_name	column_name	data_type
1	customers	customer_id	STRING
2	customers	customer_unique_id	STRING
3	customers	customer_zip_code_prefix	INT64
4	customers	customer_city	STRING
5	customers	customer_state	STRING

Insights :

1. Identifier fields are correctly stored as STRING, which helps in avoiding accidental numerical calculations, loss of formatting (e.g leading characters), Joining Inconsistencies across tables.
2. Data types are consistent, simple, and analytics-friendly, All columns use basic scalar data types (STRING, INT64). No nested, repeated, or complex fields.

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

```
SELECT
  MIN(order_purchase_timestamp) AS min_purchase_timestamp,
  MAX(order_purchase_timestamp) AS max_purchase_timestamp,
  DATE_DIFF(
      DATE(MAX(order_purchase_timestamp)),
      DATE(MIN(order_purchase_timestamp)),
      DAY
  ) AS time_range
FROM `target-sql-487006.targetsql_01.orders`;
```

Row	min_purchase_timestamp	max_purchase_timestamp	time_range
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC	773

Insights :

1. For clean YoY comparison, only 2017 is fully reliable. Since we only have complete data of year 2017,
2. Q4 is critical for retail (which it is), your conclusions could be misleading for year 2018 because of seasonal bias. since dataset stops in October 2018, you miss black Friday, Christmas season, you miss year end sales spikes

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

```sql
SELECT
    COUNT(DISTINCT a.customer_city)  AS city_count,
    COUNT(DISTINCT a.customer_state) AS state_count
FROM `target-sql-487006.targetsql_01.customers` AS a
LEFT JOIN `target-sql-487006.targetsql_01.orders` AS b
    ON a.customer_id = b.customer_id;
```

Row	City_Count	State_Count
1	4119	27

Insights :

1. Having customers in 27 states indicates national-level presence (Brazil has 27 states). This suggests the platform is not regionally limited, it has pan-country penetration.
2. 4,119 cities is a very wide geographic spread. delivery performance could vary significantly by region.

Business Recommendations :

1. Optimize Regional Logistics Infrastructure, With such wide geographic coverage, logistics efficiency becomes critical.
2. Focus on Revenue Concentration & Market Prioritization. Not all 27 states contribute equally to revenue.
3. Improve Regional Customer Personalization. Customer behavior likely varies across states (income, culture, demand patterns).

2. Indepth Exploration:

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

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

```
SELECT
    EXTRACT(YEAR  FROM order_purchase_timestamp) AS year,
    EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
    COUNT(DISTINCT order_id) AS sales
FROM `target-sql-487006.targetsql_01.orders`
GROUP BY year, month
ORDER BY year, month;
```

Row	year	month	sales
1	2016	9	4
2	2016	10	324
3	2016	12	1
4	2017	1	800
5	2017	2	1780
6	2017	3	2682

Insights :

1. There is a significant increase in orders from jan 2017 to mid-2017, which suggests strong customer acquisition phase, possibly aggressive marketing, brand awareness growth stage
2. Q4 Spike Confirms Retail Seasonality, Sharp peak around November 2017 (likely Black Friday effect).
3. In 2018, orders remain relatively stable (6000–7000 range) before decline near data-set end. Market maturity, Slowing growth rate, Transition from expansion to optimization phase, Growth is no longer exponential.

Business Recommendations :

1. Since order growth stabilized in 2018, focus on improving average order value, reducing logistics cost, increasing repeat purchase rate
2. Maximize Q4 Revenue Strategically, since November spike is very strong, increase inventory before Q4, lock supplier contracts early, increase digital marketing 4-6 weeks before peak. Maximize Annual Revenue during peak season.
3. Retention after seasonal spike, seasonal customers don't return. Work on Post-peak loyalty coupons, Email remarketing campaigns, Subscription Models.
4. To reduce season dependency, introduce mid-year promotional campaigns, launch category-specific flash sales in low months.

2.3. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)
a. 0-6 hrs: Dawn b. 7-12 hrs: Mornings c. 13-18 hrs: Afternoon d. 19-23 hrs: Night

```sql
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 day_label,

    COUNT(order_id) AS sales
FROM `target-sql-487006.targetsql_01.orders`
GROUP BY day_label
ORDER BY sales DESC;
```

Row	day_label	Sales
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Insights :

1. Afternoon contributes the highest share of total orders. This suggests that customers likely shop during lunch breaks or post-work afternoon hours.
2. Night Order are also strong, which indicates late-evening leisure time browsing behaviour, impulse purchases before sleep.

Business Recommendations :

1. Concentrate marketing during peak hours, since most orders are places in the afternoon and night, schedule push notifications around 12-3pm, run paid ads heavily between 7-10pm, launch limited-time flash sales in afternoon slots.
2. Dawn orders are very small compared to other slots. No need for aggressive marketing during this time.
3. Optimise operational readiness for peak windows.

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

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

```sql
SELECT
    b.customer_state,
    FORMAT_DATE('%Y-%m', DATE(a.order_purchase_timestamp)) AS order_month,
    COUNT(DISTINCT a.order_id) AS sales
FROM `target-sql-487006.targetsql_01.orders` AS a
LEFT JOIN `target-sql-487006.targetsql_01.customers` AS b
    ON a.customer_id = b.customer_id
GROUP BY 1,2
ORDER BY 1,2
```

Row	customer_state	order_month	sales
1	AC	2017-01	2
2	AC	2017-02	3
3	AC	2017-03	2
4	AC	2017-04	5
5	AC	2017-05	8
6	AC	2017-06	4
7	AC	2017-07	5
8	AC	2017-08	4
9	AC	2017-09	5
10	AC	2017-10	6
11	AC	2017-11	5

Insights :

1.Large scales variation across states, a few states likely SP, RJ, MG dominate sales. most states contribute small volumes. revenue concentration risk is high.
2. SP appears most frequently and has highest values, this suggests Sao Paulo is the primary revenue engine.
3. Volatility in smaller states, small states likely show high month to month volatility. Faster percentage growth but low absolute volume.

Business Recommendations :

1. Reduce revenue dependency risk, since few states drive majority sales, Invest marketing budget in Tier-2 and Tier-3 states. Offer free shipping in developing regions. Introduce state-specific campaigns.
2. Improve growth in low contributing states by geo-targeted ads, local festival campaigns, partnerships with regional sellers.
3. Manage operational cost by region, high order volume states may justify regional warehouses, logistics optimization. low volume states consolidated shipments, zonal pricing strategy.

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

```sql
SELECT
    customer_state,
    COUNT(DISTINCT customer_id) AS distribution
FROM `target-sql-487006.targetsql_01.customers`
GROUP BY customer_state
ORDER BY distribution DESC;
```

Row	customer_state	distribution
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045

Insights :

1. Heavy customer concentration in few states, SP alone contributes an extremely large share compared to others
2. SP, RJ, MG, RS, PR these 5 states dominate the business accounting for a large portion of total customers around 74-80%
3. RR, AP, AC, AM have extremely low customer counts.

Business Recommendations :

1. Since business is heavily dependent on few states diversify geographic risk, increase marketing investment in underperforming northern & northeastern states.
2. Strengthen dominant states further, top states already generate demand. So, launch premium services, increase AOV - Average Order Value, Make customers spend more per purchase.
3. Improve logistics in low-volume states, introduce zonal pricing, offer free shipping thresholds.

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

4.1. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
You can use the "payment_value" column in the payments table to get the cost of orders.

```sql
WITH tab1 AS (
    SELECT
        EXTRACT(YEAR FROM b.order_purchase_timestamp) AS payment_year,
        SUM(a.payment_value) AS year_wise_cost
    FROM `target-sql-487006.targetsql_01.payments` AS a
    JOIN `target-sql-487006.targetsql_01.orders` AS b
        ON a.order_id = b.order_id
    WHERE DATE(b.order_purchase_timestamp)
          BETWEEN '2017-01-01' AND '2018-08-31'
    GROUP BY payment_year
)

SELECT
    ROUND(MAX(CASE WHEN payment_year = 2017 THEN year_wise_cost END), 2)
        AS year_wise_cost_2017,
    ROUND(MAX(CASE WHEN payment_year = 2018 THEN year_wise_cost END), 2)
        AS year_wise_cost_2018,
    ROUND(
        (MAX(CASE WHEN payment_year = 2018 THEN year_wise_cost END)
            -
        MAX(CASE WHEN payment_year = 2017 THEN year_wise_cost END))
        /
        MAX(CASE WHEN payment_year = 2017 THEN year_wise_cost END) * 100
    , 2) AS percentage_increment
FROM tab1;
```

Row	year_wise_cost_2...	year_wise_cost_2...	percentage_incre...
1	7249746.73	8694733.84	19.93

Insights :

1. Strong Year-over-Year Growth (~20%) which indicates, higher customer acquisition, increased order frequency, higher AOV(Average Order Value)-customers spending more per purchase
2. Higher payment value means, more money flowing to sellers, more commission revenue for platform, more freight revenue for logistics partners, larger digital payment volume.
3. Revenue growth could be due to increase in number of orders, increase in AOV, expansion into more states.

Business Recommendations :

1. Since revenue is growing strongly, Increase digital marketing in high performing regions, expand catalog in top-selling categories, strengthen supplier partnerships
2. Revenue growth is strong - now improve efficiency. bundle products, introduce cart-level discounts, offer add more and save more deals, promote higher margin products
3. If revenue increases faster than order count, freight cost management becomes critical. use regional warehouses, improve delivery route optimisation.

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

```sql
WITH cte AS (
    SELECT
        a.customer_state,
        b.order_id,
        ROUND(SUM(c.price), 2) AS total_value
    FROM `target-sql-487006.targetsql_01.customers` AS a
    JOIN `target-sql-487006.targetsql_01.orders` AS b
        ON a.customer_id = b.customer_id
    JOIN `target-sql-487006.targetsql_01.order_items` AS c
        ON b.order_id = c.order_id
    GROUP BY
        a.customer_state,
        b.order_id
)

SELECT
    customer_state,
    ROUND(SUM(total_value), 2) AS total_value,
    ROUND(AVG(total_value), 2) AS avg_value
FROM cte
GROUP BY customer_state
ORDER BY avg_value DESC;
```

Row	customer_state	total_value	avg_value
1	PB	115268.08	216.67
2	AP	13474.3	198.15
3	AC	15982.95	197.32
4	AL	80314.81	195.41
5	RO	46140.64	186.8
6	PA	178947.81	184.48

Insights :

1. SP has highest total revenue but lower AOV. some smaller states (eg. PB, AC, AP) show higher AOV but lower total revenue. which indicates SP has high volume of lower value orders, some smaller states have lower volume orders but higher per-order spend.
2. States with high AOV, Low total revenue likely have low customer penetration, untapped demand, infrastructure or logistics barriers.
3. Geographic diversification is still limited, heavy revenue reliance on SP + RJ + MG, economic downturn in one major region could significantly impact total revenue.

Business Recommendations :

1. Segment strategy by state type, for volume states (SP, RJ, MG) focus on increasing AOV, bundling products, cross-sell/upsell strategies, loyalty programs. for high-AOV but low-revenue states focus on customer acquisition campaigns, logistics cost reduction.
2. Develop state level KPI dashboard, for each state track Total Revenue, AOV, Order Count, Freight %, Customer Growth %. This allows in precision decision making.
3. High AOV states may handle freight well. test dynamic shipping pricing, reduced freight promotions in key states, free shipping above high cart value.

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

```sql
WITH cte AS (
    SELECT
        a.customer_state,
        b.order_id,
        ROUND(SUM(c.freight_value), 2) AS total_freight_value
    FROM `target-sql-487006.targetsql_01.customers` AS a
    JOIN `target-sql-487006.targetsql_01.orders` AS b
        ON a.customer_id = b.customer_id
    JOIN `target-sql-487006.targetsql_01.order_items` AS c
        ON b.order_id = c.order_id
    GROUP BY
        a.customer_state,
        b.order_id
)

SELECT
    customer_state,
    ROUND(SUM(total_freight_value), 2) AS total_freight_value,
    ROUND(AVG(total_freight_value), 2) AS avg_freight_value
FROM cte
GROUP BY customer_state
ORDER BY avg_freight_value DESC;
```

Row	customer_state	total_freight_value	avg_freight_value
1	RR	2235.19	48.59
2	PB	25719.73	48.35
3	RO	11417.38	46.22
4	AC	3686.75	45.52
5	PI	21218.2	43.04

Insights :

1. States like RR, PB, RO, AC, PI Have higher average freight cost, because of geographic distance from distribution centers, lower shipment density, poor logistics infrastructure.
2. Freight efficiency should not be judged only by average, total logistics burden also matters. some states may have moderate freight per order, but high total freight cost due to high volume.
3. States with low average freight likely have, better delivery network, higher shipment density, closer warehouse proximity. These states are easier to scaler profitably.

Business Recommendations :

1. For high freight states, set up micro fulfillment centers, use third party regional warehousing, partner with local courier networks.
2. Introduce dynamic freight pricing, instead of flat shipping cost, charge region-based freight, offer free shipping above specific cart value, use weighted shipping optimization.
3. Freight reduces when volume increases, offer special campaigns in remote states. encourage bulk buying, offer "Buy More Save More".
4. Analyze freight as % of revenue, next step analysis, sate-wise freight ratio sum(freight_value)/sum(price). If freight % is too high, focus logistics improvement before aggressive scaling.

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

5.1. 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:
a. time_to_deliver = order_delivered_customer_date - order_purchase_timestamp
b. diff_estimated_delivery = order_delivered_customer_date - order_estimated_delivery_date

```sql
SELECT
    order_id,
    order_purchase_timestamp,
    order_estimated_delivery_date,
    order_delivered_customer_date,
    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-sql-487006.targetsql_01.orders`
WHERE
    order_status = 'delivered'
ORDER BY
    order_purchase_timestamp;
```

Row	order_id	order_purchase_timestamp	order_estimated_delivery_date	order_delivered_customer_date	time_to_deliver	diff_estimated_d...
1	bfbd0f9bdef84302105ad712db...	2016-09-15 12:16:38 UTC	2016-10-04 00:00:00 UTC	2016-11-09 07:47:38 UTC	54	-36
2	3b697a20d9e427646d9256791...	2016-10-03 09:44:50 UTC	2016-10-27 00:00:00 UTC	2016-10-26 14:02:13 UTC	23	0
3	be5bc2f0da14d8071e2d45451...	2016-10-03 16:56:50 UTC	2016-11-07 00:00:00 UTC	2016-10-27 18:19:38 UTC	24	10
4	a41c8759fbe7aab36ea07e038b...	2016-10-03 21:13:36 UTC	2016-11-29 00:00:00 UTC	2016-11-03 10:58:07 UTC	30	25
5	d207cc272675637bfed0062edf...	2016-10-03 22:06:03 UTC	2016-11-23 00:00:00 UTC	2016-10-31 11:07:42 UTC	27	22

Insights :

1. Freight cost correlates with remoteness. States like RR, PB, RO, AC, PI have higher freight cost. likely because of geographical distance from distribution centers, poor logistics infrastructure.
2. Freight % impact on small orders. In low AOV states, freight could represent a high percentage of order value.
3. Total freight value shows hidden cost burden, states may have moderate freight per order. but high total freight cost due to high volume.

Business Recommendations :

1. Establish regional warehousing strategy. For high freight states, set up micro fulfillment centers, use third party regional warehousing, partner with local courier networks.
2. Freight reduces when volume increases. offer special campaigns in remote states, encourage bulk buying, offer "buy more save more"
3. If freight % is too high, focus on logistics improvement before aggressive scaling.
4. If orders originate far from warehouse, consider decentralized supplier base, seller clustering strategy, state-wise vendor expansion.

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

```sql
WITH tab1 AS (
    SELECT a.customer_state, b.order_id,
        SUM(c.freight_value) AS order_freight_value
    FROM `target-sql-487006.targetsql_01.customers` a
    JOIN `target-sql-487006.targetsql_01.orders` b
        ON a.customer_id = b.customer_id
    JOIN `target-sql-487006.targetsql_01.order_items` c
        ON b.order_id = c.order_id
    GROUP BY a.customer_state,b.order_id
),
tab2 AS (
    SELECT customer_state,
    ROUND(AVG(order_freight_value), 2) AS avg_freight_value
    FROM tab1 GROUP BY customer_state)

SELECT customer_state, avg_freight_value FROM tab2
ORDER BY avg_freight_value DESC
LIMIT 5;
```

Top 5 States with Highest Average Freight Value

Row	customer_state	avg_freight_value
1	RR	48.59
2	PB	48.35
3	RO	46.22
4	AC	45.52
5	PI	43.04

```sql
SELECT customer_state, avg_freight_value FROM tab2
ORDER BY avg_freight_value
LIMIT 5;
```

Top 5 States with Lowest Average Freight Value

Row	customer_state	avg_freight_value
1	SP	17.37
2	MG	23.46
3	PR	23.58
4	DF	23.82
5	RJ	23.95

Insights :

1. There is a strong negative relationship between order volume, average freight cost. High volume states (SP, MG, RJ) - lower freight per order. Low volume states (RR, PB, R0, AC, PI) - higher freight per order.
2. High freight states (RR, RO, AC, PI) mostly north/remote regions. freight cost is likely driven by distance from warehouse, low shipment consolidation, limited carrier competition.
3. In freight states, if company subsidizes shipping, lower margins. If company passes cost to customer , lower conversion rate. This creates a strategic tradeoff.
4. Metro advantage confirmation, states like SP has logistics hubs likely located nearby, better courier competition, dense urban drop-off routes. This gives structural cost advantage.

Business Recommendations :

1. Regional micro-fulfillment strategy. for high-freight states - establish small regional stocking hubs, use third party logistics partnerships, enable cross-docking in nearby urban centers.
2. Freight subsidy strategy - subsidize freight only for high-margin categories, offer targeted free shipping campaigns.
3. Freight drops when shipment density increases, offer "buy more save more", encourage monthly subscription orders.tering strategy, state-wise vendor expansion.

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

```sql
WITH tab1 AS (
    SELECT a.customer_state,b.order_id,
        DATE_DIFF(DATE(b.order_delivered_customer_date),
            DATE(b.order_purchase_timestamp),DAY) AS delivery_duration
    FROM `target-sql-487006.targetsql_01.customers` AS a
    JOIN `target-sql-487006.targetsql_01.orders` AS b
        ON a.customer_id = b.customer_id
    WHERE b.order_status = 'delivered'
),

tab2 AS (
    SELECT customer_state,ROUND(AVG(delivery_duration), 2) AS avg_delivery_time
    FROM tab1 GROUP BY customer_state )

SELECT customer_state,avg_delivery_time FROM tab2
ORDER BY avg_delivery_time DESC
LIMIT 5;
```

Top 5 States with Highest Average Delivery Time

Row	customer_state	avg_delivery_time
1	RR	29.34
2	AP	27.18
3	AM	26.36
4	AL	24.5
5	PA	23.73

Top 5 States with Highest Average Delivery Time

```sql
SELECT customer_state,avg_delivery_time FROM tab2
ORDER BY avg_delivery_time
LIMIT 5;
```

Row	customer_state	avg_delivery_time
1	SP	8.7
2	PR	11.94
3	MG	11.94
4	DF	12.9
5	SC	14.9

Insights :

1. Strong correlation between order volume and delivery speed.
2. High delivery time states are mostly northern, remote regions, lower infrastructure development, longer last-mile distances. These are structural disadvantages.
3. If delivery takes 25-28 days(RR/AP) vs 8-12 days(SP/DF). That is a huge customer experience gap. which might effect lower repeat purchases, higher cancellation rates, lower customer lifetime value (CLV).

Business Recommendations :

1. Establish regional micro-hubs in slow states. For RR, AP, AM, PA. Use Third-party fulfillment hubs, set up small inventory pools, pre-position high-demand products. goal is to reduce delivery time from 25 to less than 15 days.
2. Use different service levels, urban metro strategy same day or next day delivery, premium logistics experience. or remote strategy cost-efficient delivery, inventory pooling.
3. Long Term Infrastructure strategy. Analyse where new warehouse placement gives maximum delivery reduction. delivery time heatmap analysis. Investment decision should be data driven.

5.4. 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.

```sql
WITH delivery_data AS (
    SELECT c.customer_state,DATE_DIFF(DATE(o.order_delivered_customer_date),
        DATE(o.order_purchase_timestamp),DAY) AS actual_delivery_days,
        DATE_DIFF(DATE(o.order_estimated_delivery_date),
        DATE(o.order_purchase_timestamp),DAY) AS estimated_delivery_days
    FROM `target-sql-487006.targetsql_01.customers` AS c
    JOIN `target-sql-487006.targetsql_01.orders` AS o
        ON c.customer_id = o.customer_id WHERE o.order_status = 'delivered')

SELECT customer_state,
    ROUND(AVG(estimated_delivery_days), 2) AS avg_estimated_days,
    ROUND(AVG(actual_delivery_days), 2) AS avg_actual_days,
    ROUND(AVG(estimated_delivery_days)-AVG(actual_delivery_days),2)
    AS delivery_time_advantage
FROM delivery_data GROUP BY customer_state
ORDER BY delivery_time_advantage DESC
LIMIT 5;
```

Row	customer_state	avg_estimated_d...	avg_actual_days	delivery_time_adv...
1	AC	41.73	21.0	20.73
2	RO	39.39	19.28	20.1
3	AP	46.87	27.18	19.69
4	AM	45.92	26.36	19.57
5	RR	46.63	29.34	17.29

Insights :

1. Top 5 fastest states compared to estimated time are low order states. which suggests estimated delivery times are being padded with a safety buffer, logistics performance is better than expected.
2. If orders are consistently delivered much earlier than estimated, customers may perceive service as slow before placing the order. Conversion rate may reduce because estimated duration looks long.
3. High volume states show estimation precision. states like SP, PR, MG has lower estimation gap, higher order density, stronger route planning accuracy.

Business Recommendations :

1. Improve delivery time prediction model using historical delivery time, region level transit time, order density, warehouse proximity. Goal is to reduce gap between estimated and actual delivery time.
2. Reduce over buffering in remote states. If delivery consistently beats estimate by several days, reduce estimated delivery window, improve displayed ETA confidence. Which impact higher conversion rate.
3. Analyze delivery time vs repeat purchase. Test are early deliveries linked to higher repeat orders, if yes then ETA optimization becomes revenue lever.

6. Analysis based on the payments.

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

```sql
SELECT
    FORMAT_DATE('%Y-%m', DATE(a.order_purchase_timestamp)) AS order_month,
    b.payment_type,COUNT(DISTINCT a.order_id) AS total_orders
FROM `target-sql-487006.targetsql_01.orders` AS a
JOIN `target-sql-487006.targetsql_01.payments` AS b
    ON a.order_id = b.order_id
GROUP BY
    order_month,b.payment_type
ORDER BY
    order_month,b.payment_type DESC,total_orders DESC;
```

Row	order_month	payment_type	total_orders
1	2016-09	credit_card	3
2	2016-10	voucher	11
3	2016-10	debit_card	2
4	2016-10	credit_card	253
5	2016-10	UPI	63
6	2016-12	credit_card	1
7	2017-01	voucher	33
8	2017-01	debit_card	9
9	2017-01	credit_card	582

Insights :

1. Credit cards consistently show the highest order volume. This indicates higher trust in digital payments, urban/financially active user base, possibly higher order value (credit usage allows bigger purchases).
2. UPI shows strong upward trend month-over-month. This signals fast adoption of instant payment systems. Younger/mobile first customers. Preference for faster checkout experiences.
3. Voucher payments spike in specific months, This suggests promotion driven behaviour. Discount campaigns influence payment choice. Customers are responsive to incentives.
4. Debit cards appears steady but not dominant.
5. Having multiple strong payment types reduces operational risk, protects from payment gateway failure, improves conversion rate.

Business Recommendations :

1. For credit card users, offer EMI options, reward points, premium product bundles, bank partnerships, bank partnerships. For UPI users cashback incentives, fast checkout UX optimization, one click payments.
2. Since UPI users are convenience driven. Loyalty points for repeat UPI use. Gamified reward system, Subscription options.
3.Smart voucher campaign optimisation. Analyze ROI of voucher based orders. compare AOV of voucher vs non-voucher orders, Avoid over-subsidizing margins.
4. With credit card dominance. Strengthen fraud monitoring, Detect unusual transaction spikes, use risk scoring. Protects long-term operational health.

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

```sql
SELECT
    payment_installments,
    COUNT(DISTINCT order_id) AS number_of_orders
FROM `target-sql-487006.targetsql_01.payments`
WHERE payment_installments >= 1
GROUP BY payment_installments
ORDER BY number_of_orders DESC;
```

Row	payment_installm...	number_of_orders
1	1	49060
2	2	12389
3	3	10443
4	4	7088
5	10	5315
6	5	5234
7	8	4253
8	6	3916
9	7	1623
10	9	644

Insights :

1. Strong preference for full payment (1 Installment). Most of the customers prefer immediate full payment, lower reliance on EMI, possible lower average order value (Impulse or small ticket purchases)
2. Sharp drop after 2-3 Installments, this suggests that customers are comfortable with short term EMIs, Long tenure adoption is weak, Possibly fewer high-ticket items.
3. Interestingly, 10-installment orders are higher than some mid-range installments, this may include certain expensive categories (electronics, appliances). Specific bank EMI promotions, Seasonal high-ticket sales.
4. Long term installment plans are low. Low usage may mean interest costs discourage customers, product basket not high value, customers avoid debt like commitments.

Business Recommendations :

1. Boost high ticket product sales via EMI marketing, since long EMIs are under utilized. Promote zero-cost EMI options, Highlight "No extra cost" messaging.
2. Optimize short-term EMI, offer cashbacks on 3-installment plans, promote "Pay in 3" offers. Partner with banks for low-interestn3-month EMI,
4. Convert single installment buyers to premium buyers. Since majority pays in full cross sell premium variants, bundle accessories, recommend complementary products.
5. Installment based customer segmentation. segment customers into immediate buyers (1 installment), short EMI users (2-4), Long EMI users(5+). Then analyze AOV difference, repeat rate, category preference, personalize offers accordingly.