Operational Analysis of Target's Brazil Operations: A Practical Learning Case Study (2016-2018)

Prepared by:

Jagan Mohan Mathangi
MSc Financial Technology
University of Nottingham

February 2025

Important: This case study is conducted solely for educational purposes and does not reflect the actual operations or financial performance of Target Corporation. The findings and insights derived from the analysis are based on hypothetical scenarios and should not be interpreted as factual representations of the company.

Content

Executive Summary

Overview:

This analysis explores Target's e-commerce operations in Brazil using data from 100,000 orders placed between September 2016 and October 2018. The dataset provides insights into customer behaviour, order trends, pricing, payment methods, shipping efficiency, and regional distribution. The findings highlighted growth opportunities, operational efficiencies, and strategic recommendations to optimize performance.

Key Findings & Insights:

1. **Market Growth & Order Trends:**
    a. **Strong Growth:** Orders increased significantly, with a 20% YoY growth in 2018.
    b. **Seasonal Patterns:** May-August saw peak order volumes, with August recording the highest volume, and September-December had the lowest volume.
    c. **Time-of-Day Behaviour:** Afternoon and night are the most preferred shopping times.
2. **Geographic Expansion & Customer Distribution**
    a. **Broad Reach:** Orders came from 4,119 cities across all 27 states. However, 1,300 municipalities remain untapped, signalling room for expansion.
    b. **High-Concentration Regions:** São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG) account for 65% of customers.
    c. **Growth Opportunities:** Underpenetrated states need targeted marketing and localized logistics strategies.
3. **Financial & Pricing Analysis**
    a. **Massive Revenue Growth:** Recorded 136.98% YoY growth in total order value from 2017 to 2018.
    b. **Order Value by State:**
        i. SP has the highest total revenue but the lowest average order value, suggesting frequent low-value purchases.
        ii. Paraíba (PB) has the highest average order price ($216.67), indicating potential for premium sales.
    c. **Freight Costs:**
        i. SP has the lowest average freight ($15.15) due to its strong logistics infrastructure.
        ii. Roraima (RR) has the highest average freight ($42.98), making affordability a challenge.
4. **Delivery Efficiency & Logistics**
    a. **Delivery Times:** Orders took between 29-67 days for delivery, with significant delays in remote regions.
    b. **Fastest & Slowest Delivery States:**
        i. SP has the fastest delivery time (8.3 days) due to its infrastructure.
        ii. RR (28.98 days) and Amazon regions suffer long delays, requiring logistics improvements.
    c. **Early Deliveries:** Acre (AC) delivers orders 20 days earlier than estimated, suggesting room for improved delivery forecasting.

5. **Payment Trends & Customer Behaviour**
    a. **Credit Cards Dominate:** This is the most preferred payment method across all months.
    b. Installment Payments:
        i. Single-payment orders are common, indicating strong customer financial stability.
        ii. 99.6% of installment-based orders use 10-month financing, highlighting a strong market for long-term payment plans.

Strategic Recommendations:

1. **Expand Into Untapped Markets**
    a. Target the 1,300 uncovered municipalities with localized advertising and logistical improvements.
    b. Offer regional promotions to boost engagement in underperforming states.
2. **Optimize Seasonal Demand**
    a. Introduce marketing pushes and discounts from September-December to counteract seasonal slumps.
    b. Leverage holiday promotions and loyalty programs during low-demand periods.
3. **Improve Logistics in High-Cost/Slow Regions**
    a. Set up regional fulfillment centres to cut freight costs in RR, PB, and remote Amazonian states.
    b. Partner with local delivery services to reduce delays and shipping costs.
4. **Leverage Payment Trends for Increased Sales**
    a. Promote installment-based pricing to align with the high demand for 10-month financing.
    b. Offer UPI-based incentives to drive adoption of digital payments.
5. **Enhance Delivery Time Predictions**
    a. Improve estimated delivery time accuracy, particularly in Acre (AC) and other early-delivery states.
    b. Use real-time tracking and AI-based forecasting to reduce delays in slower regions.

Conclusion:

Target Brazil has experienced strong growth, increasing customer demand, and financial success over the past two 2017 and 2018. However, untapped markets, high freight costs, and delivery inefficiencies in remote areas present challenges. By implementing targeted expansion strategies, optimizing logistics, and leveraging financial insights, Target can further strengthen its e-commerce dominance in Brazil.

Company Introduction

- Target Corporation provides daily essentials and fashionable, differentiated merchandise to its customers at discounted prices.
- Operates through a single segment, allowing customers to purchase products either in stores or through digital channels.
    - The digital platform offers a broad assortment of merchandise and food products, including complementary selections sold by third-party vendors.
- Manages its business under five merchandise categories
    - Apparel & Accessories
    - Beauty & Household Essentials
    - Food & Beverage
    - Hardlines
    - Home Furnishings & Decor
- Owned and exclusive brands account for 33% of the total revenue
    - Owned portfolio includes A New Day, All in Motion, Boots & Barkley, Casaluna, Cloud Island, Embark, Made By Design, More Than Magic, Smith & Hawken, Original Use, among others.
    - Additionally, offers exclusive adult beverage brands such as California Roots, Jingle & Mingle, Rosé Bae, Casa Cantina, The Collection, SunPop, Photograph, and others.
- Distributes most of its merchandise to its stores through a network of distribution centres.
    - Offers same-day delivery through its wholly owned subsidiary, **Shipt**
    - Manages 62 supply chain facilities across the U.S.
    - Operates over 20 sourcing offices worldwide
- Competes with traditional and online retailers, including department stores, off-price general merchandise retailers, wholesale clubs, category-specific retailers, drug stores, supermarkets, direct-to-consumer brands, and other forms of retail commerce.
- As of February 2024, Target operates 1,956 stores, which include:
    - 1,532 company – owned stores
    - 264 leased stores
    - 160 stores with owned buildings on leased land
- Employs over 415,000 people across the U.S. and India
- Founded in 1902 and headquartered in Minneapolis, MN

List of Questions

Section	Questions
1. Initial Data Exploration	1. Data Types in the Customers Table: Identify the data types of all columns in the "customers" table. 2. Order Time Range: Determine the time range within which the orders were placed. 3. City and State Counts: Count the number of unique cities and states from which customers placed orders during the specified period.
2. In-depth Exploration	4. Trend Analysis: Investigate whether there is a growing trend in the number of orders placed over the past years. 5. Monthly Seasonality: Analyse if there is any monthly seasonality in the number of orders being placed. 6. Order Timing Analysis: Determine the time of day when Brazilian customers primarily place their orders, categorizing them as follows: • 0-6 hrs: Dawn • 7-12 hrs: Mornings • 13-18 hrs: Afternoon • 19-00 hrs: Night
3. Evolution of E-commerce Orders	7. Monthly Orders by State: Retrieve the month-on-month number of orders placed in each state. 8. Customer Distribution: Assess how customers are distributed across all states.
4. Economic Impact Analysis	9. Cost Increase Calculation: Analyse the percentage increase in the cost of orders from 2017 to 2018, including only the months from January to August. Use the "payment_value" column from the payments table for this analysis. 10. Order Price Analysis by State: Calculate the total and average order price for each state. 11. Freight Value Analysis by State: Calculate the total and average freight value for each state.

5. Sales, Freight, and Delivery Time Analysis	12. Delivery Time Calculation: Find the number of days taken to deliver each order from the order's purchase date. Additionally, calculate the difference (in days) between the estimated and actual delivery dates of each order, using the following formulas: • Delivery Time: order_delivered_customer_date - order_purchase_timestamp • Difference in Estimated Delivery: order_delivered_customer_date - order_estimated_delivery_date 13. Freight Value Rankings: Identify the top five states with the highest and lowest average freight values. 14. Delivery Time Rankings: Identify the top five states with the highest and lowest average delivery times. 15. Fastest Delivery Rankings: Identify the top five states where order delivery is significantly faster than the estimated delivery date, using the difference between the averages of actual and estimated delivery dates.
6. Payment Analysis	16. Monthly Payment Type Analysis: Find the month-on-month number of orders placed using different payment types. 17. Payment Instalment Analysis: Determine the number of orders placed based on the payment instalments made.

Data Overview

Data Sources:

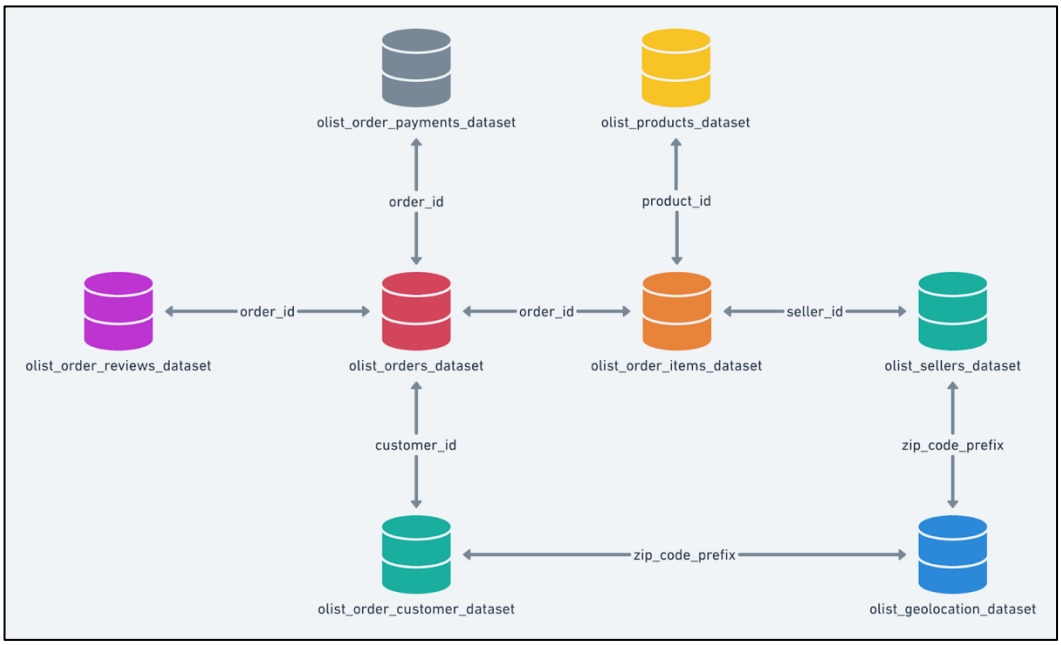
The dataset is sourced from Kaggle. It consists of eight CSV files, representing different aspects of an e-commerce platform, such as customers, sellers, orders, payments, reviews, and geolocation data.

Data Structure:

The dataset contains the following tables:

- **customers.csv** – Contains details about customers, including unique customer IDs, location information (city, state, zip code).
- **sellers.csv** – Stores information about registered sellers, including their ID, location details, and zip code prefix.
- **order_items.csv** – Represents items within an order, tracking the order ID, product ID, seller ID, price, and shipping details.
- **geolocation.csv** – Provides geographical coordinates (latitude, longitude) mapped to zip codes, cities, and states.
- **payments.csv** – Stores payment details, including payment type, installment count, and total amount paid per order.
- **orders.csv** – Tracks orders with details such as status, timestamps, and estimated delivery dates.
- **reviews.csv** – Contains customer reviews for orders, including scores, comments, and timestamps.
- **products.csv** – Lists product details, such as category, dimensions, and weight.

Dataset Schema:



Data Analysis and Results:

Initial Data Exploration

Question – 1: Identify the data types of all columns in the "customers" table.

To identify the data types of all columns in the customers table, I executed the following query:

```
SELECT
      column_name, data_type
FROM
      `target.information_scheme.columns`
WHERE
      table_name = 'customers'
```

The query selects the `column_name` and `data_type` from the `information_schema.columns` view, filtering for the table named customers. The results, as shown in Figure 1.1.1, indicate that the customers table contains five columns: `customer_id, customer_unique_id, customer_zip_code_prefix , customer_city,` and `customer_state`.

- The columns `customer_id, customer_unique_id, customer_city`, and `customer_state` have a `STRING` data type, which means they store text-based information.
- The column `customer_zip_code_prefix` has an `INT64` data type, indicating that it stores integer values.

Figure 1.1: SQL Query Result: Data Types of Columns in "customers" Table

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

Question – 2: Determine the time range within which the orders were placed.

To determine the time range within which the orders were placed, I executed the following query:

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

This query retrieves the earliest (`MIN`) and latest (`MAX`) order dates from the orders table, effectively providing the time range for all orders in the dataset. The results, as shown in Figure 1.2.1, indicate that the first order was placed on September 4, 2016, at 21:15:19 UTC, and the last order was placed on October 17, 2018, at 17:30:18 UTC.

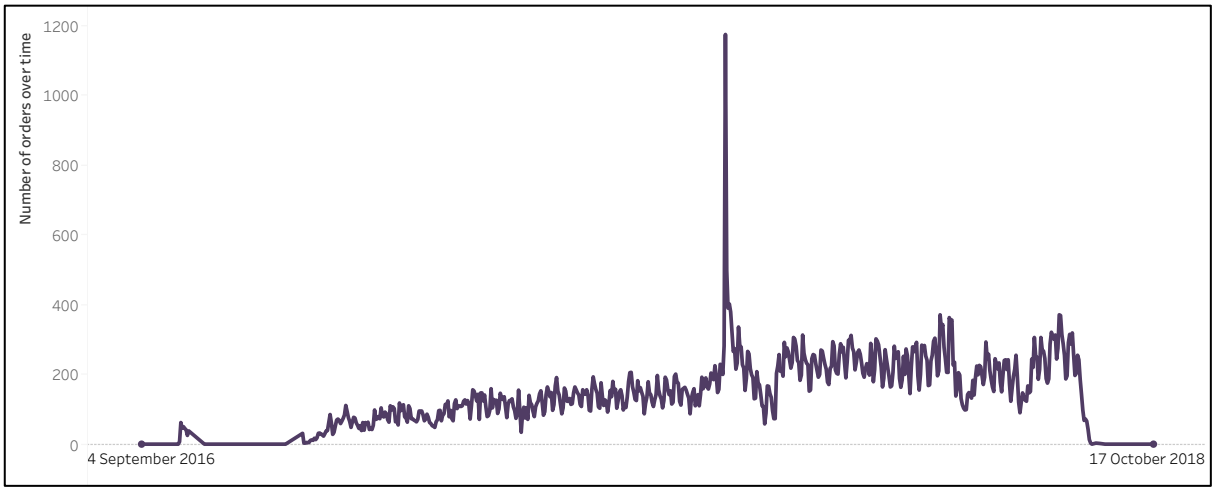
Figure 2.1: Order Date Range (Query Output)

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

Key Insights:

- The dataset covers a time period of approximately 2 years and 1 month.
- The first order was placed in September 2016, and the last order was placed in October 2018.

Figure 2.2: Numbers of Orders Over Time



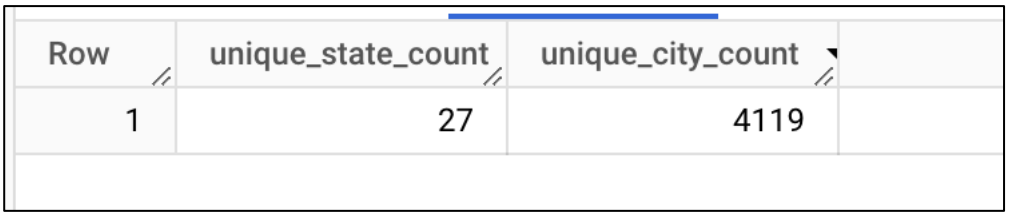
Question – 3: Count the number of unique cities and states from which customers placed orders during the specified period.

To determine the total number of unique cities and states from which the Brazilian customers placed orders within the two-year and one-month time span, I executed the following query:

```
SELECT
COUNT(DISTINCT(customer_state)) AS unique_state_count,
COUNT(DISTINCT(customer_city)) AS unique_city_count
FROM
`target.customers`
```

This query calculates the number of unique states (`customer_state`) and cities (`customer_city`) in the customers table. The results, as shown in Figure 1.3.1, indicate that Target received orders from 4,119 unique cities across 27 states (including the Federal District).

Figure 3.1: Number of Unique States and Cities (Query Output)

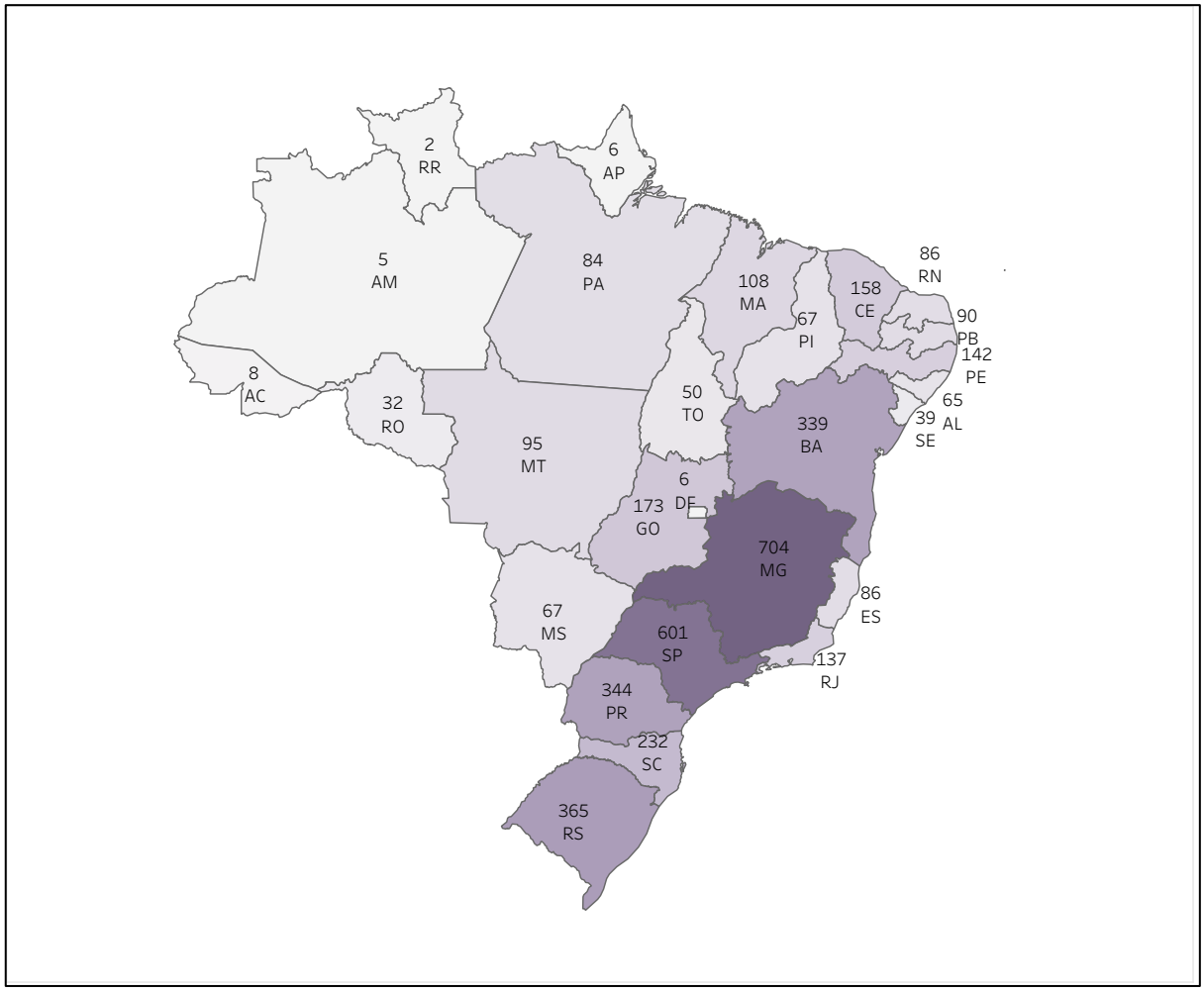


Row	unique_state_count	unique_city_count
1	27	4119

Key Insights:

- **Geographical Reach:** Target has a significant presence in Brazil, with orders coming from 4,119 cities out of over 5,500 recognized municipalities.
- **Untapped Potential:** There are over 1,300 municipalities where Target has not yet received orders, representing a substantial opportunity for expansion.
- **State Coverage:** Orders were received from all 27 states, including the Federal District, indicating a widespread customer base.

Figure 3.2: Number of Unique Cities by State



In-depth Exploration

Question – 4: Investigate whether there is a growing trend in the number of orders placed over the past years.

To understand whether there is a growing trend in the number of orders placed over the past years, I executed the following query:

```
SELECT
  EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
  COUNT(order_purchase_timestamp) AS no_of_orders
FROM `target.orders`
GROUP BY year
ORDER BY no_of_orders
```

This query extracts the year from the `order_purchase_timestamp` and counts the number of orders placed each year. The results, as shown in Figure 1.4.1, reveal that the number of orders increased significantly from 329 in 2016 to 45,101 in 2017, representing a rapid growth phase. This growth continued in 2018, with the number of orders reaching 54,011, a ~20% increase over the previous year.

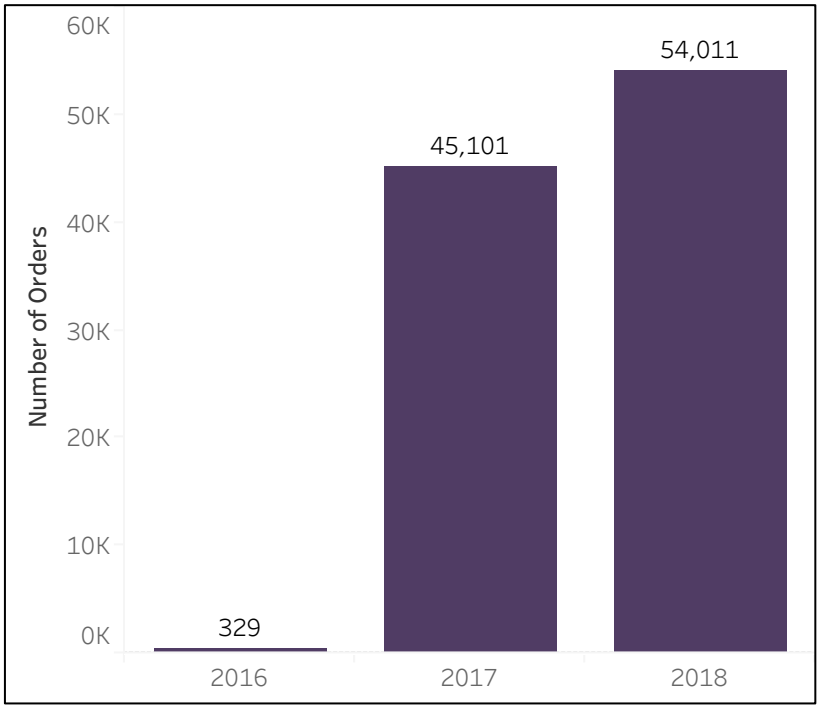
Figure 4.1: Number of Orders over Years (Query Output)

Row	year	no_of_orders
1	2016	329
2	2017	45101
3	2018	54011

Key Insights:

- **Rapid Growth in 2017:** The number of orders increased significantly in 2017 compared to 2016, indicating a high return on marketing spend and successful business strategies.
- **Steady Growth in 2018:** The company achieved a ~20% increase in orders in 2018 compared to 2017, demonstrating continued growth and customer demand.
- **Overall Trend:** The data shows a clear upward trend in the number of orders over the three-year period, reflecting Target's growing presence and success in the Brazilian e-commerce market.

Figure 4.2: Yearly Order Volume



Question – 5: Analyse if there is any monthly seasonality in the number of orders being placed.

To analyse whether there is any monthly seasonality in the number of orders placed, I executed the following query:

```
WITH monthly_seasonality AS (
  SELECT
    EXTRACT(MONTH from order_purchase_timestamp) AS Num_month,
    FORMAT_TIMESTAMP('%B', order_purchase_timestamp) AS month,
    COUNT(order_purchase_timestamp) AS no_of_orders
  FROM
    `target.orders`
  GROUP BY
    month, Num_month
  ORDER BY
    Num_month
)
SELECT
  month,
  no_of_orders
FROM monthly_seasonality
```

This query extracts the month from the `order_purchase_timestamp`, formats it as the full month name, and counts the number of orders placed each month. The results, as shown in Figure 1.5.1, reveal the monthly distribution of orders over the two-year and one-month time span.

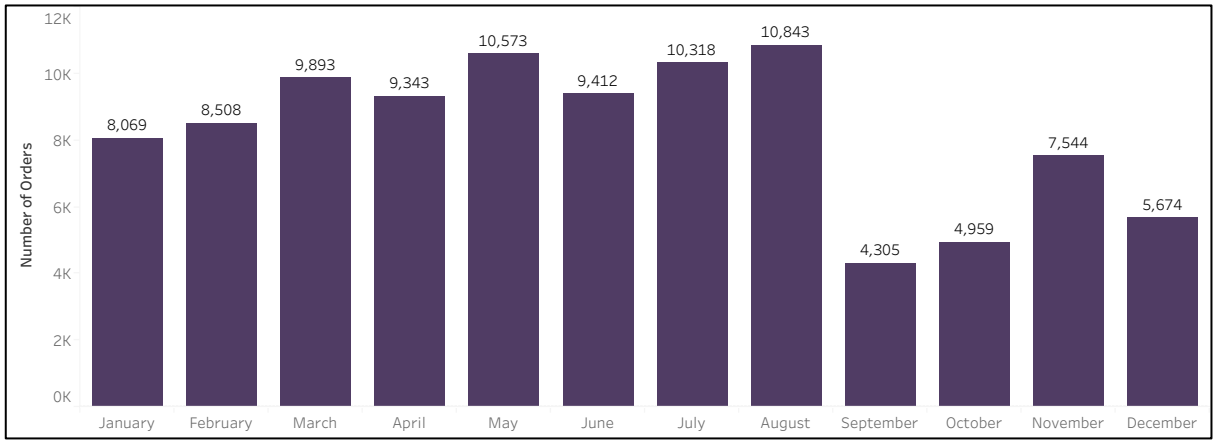
Figure 5.1: Distribution of Orders by Month (Query Output)

Row	month	no_of_orders
1	January	8069
2	February	8508
3	March	9893
4	April	9343
5	May	10573
6	June	9412
7	July	10318
8	August	10843
9	September	4305
10	October	4959
11	November	7544
12	December	5674

Key Insights:

- **Peak Months:** The company experienced a high number of orders during May, July, and August, with August being the peak month (10,843 orders).
- **Low Months:** The number of orders dropped significantly from September to December, with September having the lowest number of orders (4,305).
- **Trend:**
    - There is a clear upward trend from January to August, indicating increasing order volumes during the first eight months of the year.
    - From September to December, the company experienced a flat trend, with order volumes remaining relatively flat.

Figure 5.2: Monthly Order Distribution



Question – 6: Determine the time of day when Brazilian customers primarily place their orders, categorizing them as follows:

- **0-6 hrs: Dawn**
- **6-12 hrs: Mornings**
- **12-18 hrs: Afternoon**
- **18-00 hrs: Night**

To determine the time of day during which Brazilian customers primarily shop, I used the following query:

```
SELECT
    CASE
        WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 0 AND
EXTRACT(HOUR FROM order_purchase_timestamp) < 6 THEN 'Dawn'
        WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 6 AND
EXTRACT(HOUR FROM order_purchase_timestamp) < 12 THEN 'Morning'
        WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 12 AND
EXTRACT(HOUR FROM order_purchase_timestamp) < 18 THEN 'Afternoon'
        WHEN EXTRACT(HOUR FROM order_purchase_timestamp) >= 18 AND
EXTRACT(HOUR FROM order_purchase_timestamp) <= 23 THEN 'Night'
    END AS time_of_the_day,
    COUNT(*) AS no_of_orders
FROM
    `target.orders`
GROUP BY
    time_of_the_day
ORDER BY
    time_of_the_day;
```

This query extracts the time-of-day information from the `order_purchase_timestamp` column in the orders table. The results, as shown in the figure, reveal the distribution of orders across different times of day over the two-year and one-month period, along with the corresponding number of orders for each period.

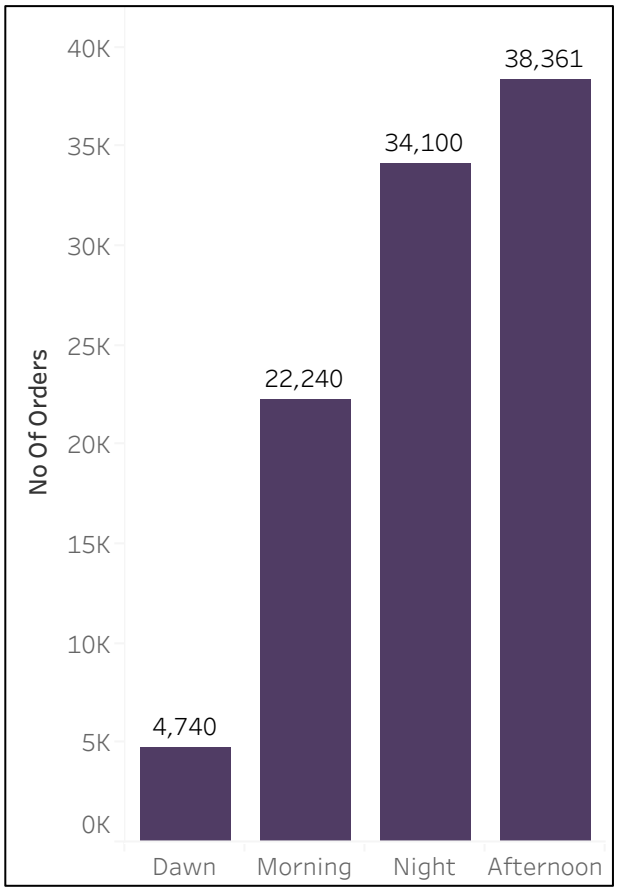
Figure 6.1: Order Distribution by Time of Day (Query Output)

Row	time_of_the_day	no_of_orders
1	Afternoon	38361
2	Dawn	4740
3	Morning	22240
4	Night	34100

Key insights

- Afternoon and night are the most preferred times for the Brazilian customers for placing their orders.
- Dawn is the least preferred time for ordering.
- Morning also sees a significant number of orders, although comparatively fewer than afternoon and night.

Figure 6.2: Order Frequency by Time of Day



Question – 7: Retrieve the month-on-month number of orders placed in each state.

To retrieve the month-on-month number of orders placed in each state, I executed the following query:

```
SELECT
  C.customer_state as state,
  EXTRACT (MONTH FROM O.order_purchase_timestamp) AS month,
  COUNT(DISTINCT(order_id)) AS total_orders
FROM
  `target.customers` AS C
  INNER JOIN `target.orders` AS O
  ON C.customer_id = O.customer_id
GROUP BY
  state, month
ORDER BY
  state, month
```

The query extracts the `state, month` and total number of orders from the orders and customers tables. The results are grouped by state and month, providing a detailed breakdown of orders across different regions and time periods. The output is visualized in Figure 1.1.1 using a bar chart.

Figure 7.1: State-level monthly order distribution

Row	state	month	total_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
11	AC	11	5
12	AC	12	5

Key Insights:

- **High-Order States:** The company received over 1,000 orders per month from only three states (e.g., São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG)), indicating strong market presence in these regions.
- **Low-Order States:** 22 states accounted for fewer than 500 orders per month, highlighting significant growth opportunities in these underpenetrated regions. This indicates a need for targeted marketing and potentially logistical improvements in these underperforming states.
- **Regional Trends:** The data reveals varying order volumes across states, with some states showing consistent performance while others exhibit seasonal fluctuations. This difference in

order volume between states suggests the need for regional marketing and operational strategies. For example, states with low order volumes might benefit from localized advertising, partnerships with local influencers, or tailored promotions.

Figure 7.2: State-level monthly order distribution

	Month											
State	1	2	3	4	5	6	7	8	9	10	11	12
AC	8	6	4	9	10	7	9	7	5	6	5	5
AL	39	39	40	51	46	34	40	34	20	30	26	14
AM	12	16	14	19	19	8	23	9	9	3	10	6
AP	11	4	8	5	11	4	7	5	2	3	4	4
BA	264	273	340	318	368	307	405	323	170	170	250	192
CE	99	101	126	143	136	121	140	130	77	74	108	81
DF	151	196	207	183	208	220	243	232	97	104	168	131
ES	159	186	182	188	228	204	206	200	93	104	170	113
GO	164	176	199	177	226	184	192	213	88	117	157	127
MA	66	67	77	73	65	59	79	70	42	52	56	41
MG	971	1,063	1,237	1,061	1,190	1,080	1,111	1,177	511	600	943	691
MS	71	75	79	58	74	76	74	59	33	34	46	36
MT	96	84	71	92	104	83	85	78	35	55	74	50
PA	82	83	109	107	75	92	96	104	41	58	70	58
PB	33	47	55	51	47	51	79	46	29	31	30	37
PE	113	146	153	154	174	140	210	170	76	87	126	103
PI	55	46	48	50	56	43	52	43	23	25	31	23
PR	443	460	504	500	524	478	523	556	183	225	378	271
RJ	990	1,176	1,302	1,172	1,321	1,128	1,288	1,307	612	725	1,048	783
RN	51	31	52	42	39	49	56	40	24	27	44	30
RO	23	25	29	20	26	22	27	23	16	14	17	11
RR	2	7	8	4	3	8	6		2	4	2	
RS	427	473	569	488	559	526	565	599	279	276	422	283
SC	345	316	362	351	379	321	356	365	157	189	303	193
SE	24	27	43	27	19	37	42	43	16	25	27	20
SP	3,351	3,357	4,047	3,967	4,632	4,104	4,381	4,982	1,648	1,908	3,012	2,357
TO	19	28	28	33	34	26	23	28	17	13	17	14
Grand Total	8,069	8,508	9,893	9,343	10,573	9,412	10,318	10,843	4,305	4,959	7,544	5,674

Question – 8: Assess how customers are distributed across all states.

To assess the customer base across all the states, I executed the following query:

```
SELECT
  customer_state AS state,
  COUNT(customer_id) AS total_customers
FROM
  `target.customers`
GROUP BY
  state
ORDER BY
  total_customers DESC
```

The query extracts the state from the customers table and counts the total customers residing in each state.

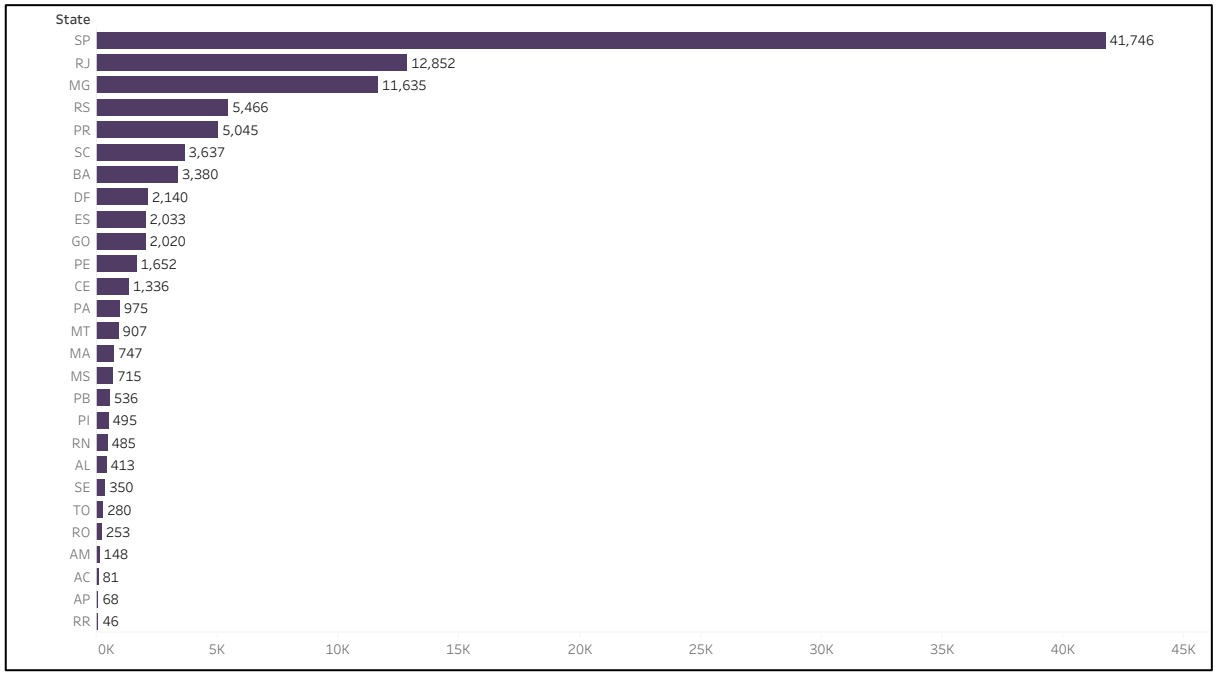
Figure 8.1: Customer Distribution across States (Query Output)

Row	state	total_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
11	PE	1652
12	CE	1336
13	PA	975

Key Insights:

- **Highly Concentrated Customer Base:** The distribution of customers is heavily skewed. A significant portion of the customer base is concentrated in just three states: SP, RJ, and MG. These three states alone account for over 65,000 customers, representing a substantial majority of the total customer base.
- **SP Dominance:** The state of SP (São Paulo) stands out with over 41,000 customers, significantly more than any other state. This suggests a strong market presence and penetration in this particular region.
- **Untapped potential:** Beyond the top three states, there's a long tail of states with significantly fewer customers. This indicates a large untapped market and potential for growth in these regions, which might require regional marketing and operational strategies as mentioned in question above.

Figure 8.2: Customer Distribution across States



Economic Impact Analysis

Question – 9: Analyse the percentage increase in the cost of orders from 2017 to 2018, including only the months from January to August. Use the "payment_value" column from the payments table for this analysis.

To calculate the percentage increase in the cost of orders from 2017 to 2018, considering only January to August, I executed the following query:

```
WITH CTE AS (
  SELECT
  1 AS key,
  EXTRACT (YEAR FROM O.order_purchase_timestamp) AS year,
  ROUND(SUM(payment_value),2) AS Y2017
FROM `target.payments` AS P INNER JOIN `target.orders` AS O ON
P.order_id = O.order_id
WHERE EXTRACT(MONTH FROM order_purchase_timestamp) BETWEEN 1 AND 8
GROUP BY year
HAVING year = 2017
),
CTE1 AS (
  SELECT
  1 AS key,
  EXTRACT (YEAR FROM O.order_purchase_timestamp) AS year,
  ROUND(SUM(payment_value),2) AS Y2018
FROM `target.payments` AS P INNER JOIN `target.orders` AS O ON
P.order_id = O.order_id
WHERE EXTRACT(MONTH FROM order_purchase_timestamp) BETWEEN 1 AND 8
GROUP BY year
HAVING year = 2018
)

SELECT
  Y2017, Y2018,
  ROUND((Y2018/Y2017-1)*100,2) AS Percentage_change
FROM CTE INNER JOIN CTE1 ON CTE.key = CTE1.key
```

This query calculates the total cost of orders (`payment_value`) for the months from January to August in 2017 and 2018 and then computes the percentage increase from 2017 to 2018. The results are shown in Figure 2.1.1.

Figure 9.1: Percentage Change in Order Cost (Query Output)

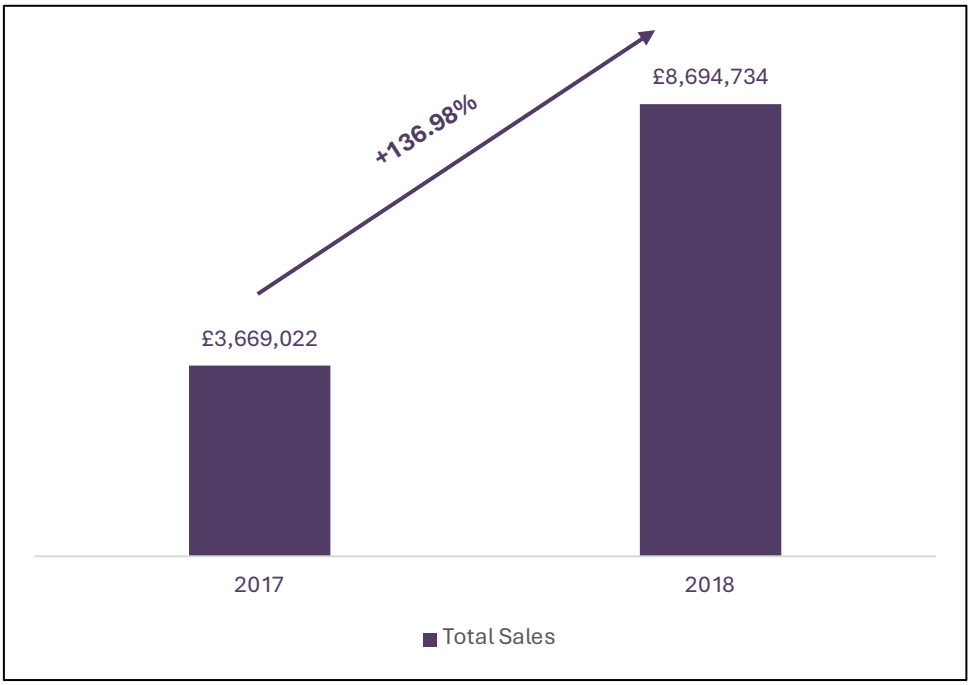
Row	Y2017	Y2018	Percentage_change
1	3669022.12	8694733.84	136.98

Key Insights:

- **Massive Growth:** The cost of orders increased by 136.98% from 2017 to 2018, highlighting the rapid growth and acceptance of Target in the Brazilian market.
- **Strong Performance:** This significant year-on-year (YoY) growth indicates successful business strategies, effective marketing campaigns, and increasing customer trust in the brand.

- **Market Potential:** The data suggests that Target has substantial potential for further growth in the Brazilian e-commerce market.

Figure 9.2: Percentage Change in Order Cost (Query Output)



Question – 10: Calculate the total and average order price for each state. (Average price per order)

To calculate the total and average order price for each state, I executed the following query:

```
SELECT
  DISTINCT(customer_state) AS state,
  ROUND(SUM(price),2) AS total_order_price,
  ROUND(SUM(price)/count(DISTINCT(OI.order_id)),2) AS
average_order_price
FROM `target.customers`
AS C INNER JOIN `target.orders` AS O ON C.customer_id = O.customer_id
INNER JOIN `target.order_items` AS OI ON OI.order_id = O.order_id
GROUP BY state
ORDER BY state
```

The query retrieves the states from customers table and calculates the total order price and average order price per state.

Figure 10.1: Total and Average Order Price by State (Query Output)

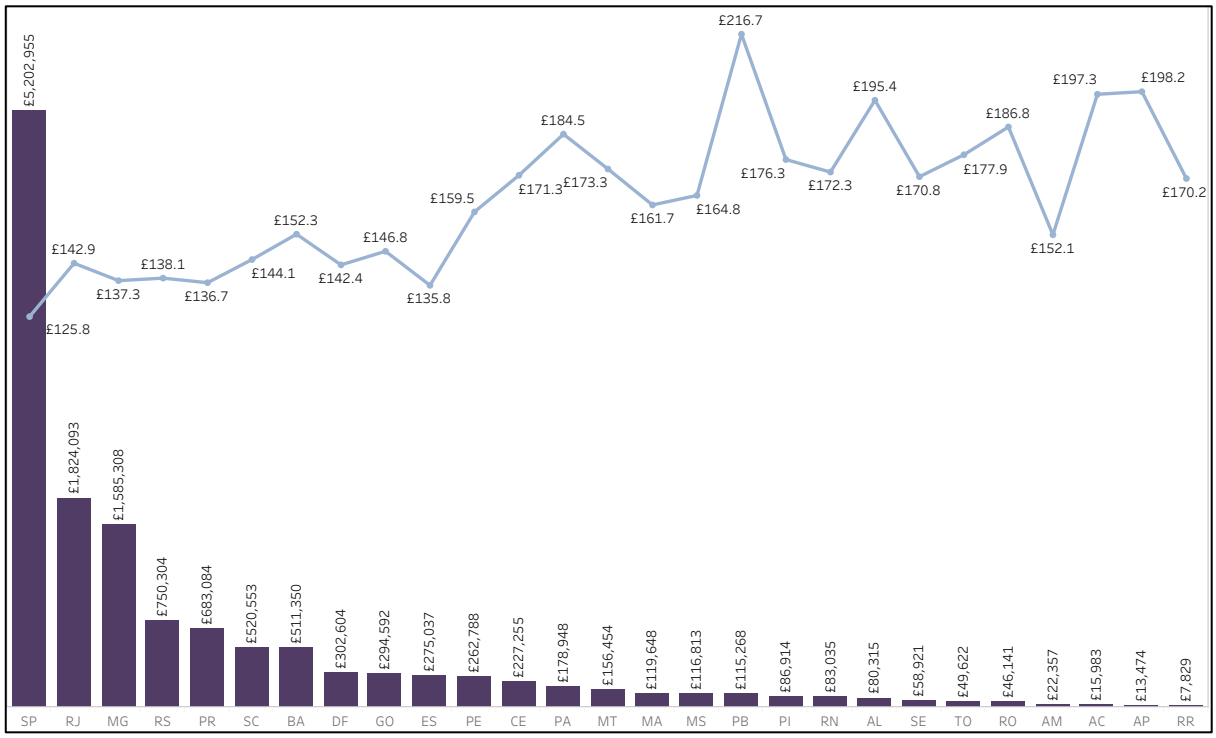
Row	state	total_order_price	average_order_price
1	AC	15982.95	197.32
2	AL	80314.81	195.41
3	AM	22356.84	152.09
4	AP	13474.3	198.15
5	BA	511349.99	152.28
6	CE	227254.71	171.25
7	DF	302603.94	142.4
8	ES	275037.31	135.82
9	GO	294591.95	146.78
10	MA	119648.22	161.69
11	MG	1585308.03	137.33
12	MS	116812.64	164.76
13	MT	156453.53	173.26

Key Insights:

- SP has the highest total order value, indicating a high number of orders from this state. However, it also has the lowest average order value. This could suggest:
  - Customers frequently order small-value products, showing strong reliance on Target.
  - Target may not have established enough trust for customers to make high-value purchases.
- PB has the highest average order price at $216.67, suggesting that customers in this state tend to purchase higher-value items.

- Fewer than 10 states have an average order value below $150. The company should focus on increasing the average order price in these states, as a $150 order value may not be sustainable when factoring in delivery costs.

Figure 10.2: Total and Average Order Price by State



Question – 11: Calculate the total and average freight value for each state.

To calculate the total and freight value for each state, I executed the following query:

```
SELECT
  DISTINCT(customer_state) AS state,
  ROUND(SUM(freight_value),2) AS total,
  ROUND(SUM(freight_value)/COUNT(DISTINCT(OI.order_id)),2) AS
average_order_price
FROM `target.customers`
AS C INNER JOIN `target.orders` AS O ON C.customer_id = O.customer_id
INNER JOIN `target.order_items` AS OI ON OI.order_id = O.order_id
GROUP BY state
ORDER BY state
```

The query retrieves the states from customers table and calculates the total freight price and average freight value per state.

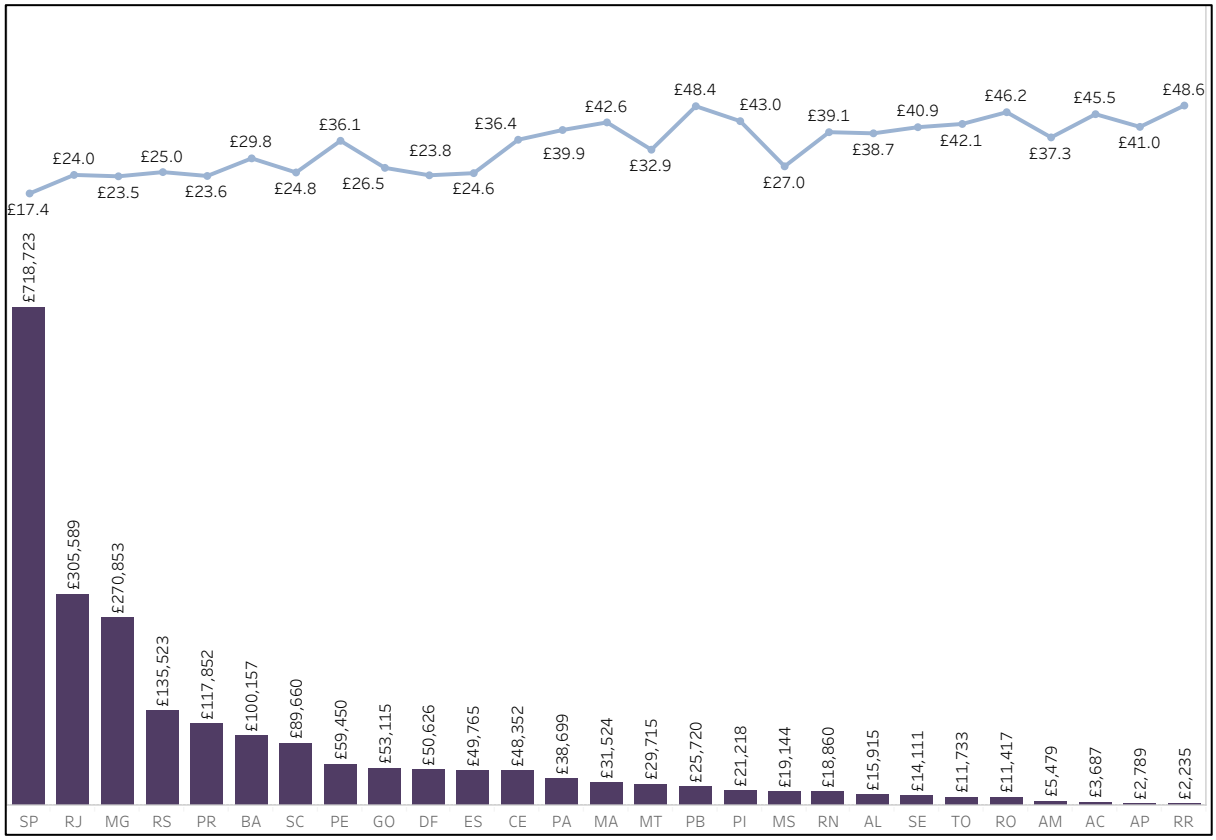
Figure 11.1: Total and Average Freight Price by State (Query Output)

Row	state	total	average_order_price
1	AC	3686.75	45.52
2	AL	15914.59	38.72
3	AM	5478.89	37.27
4	AP	2788.5	41.01
5	BA	100156.68	29.83
6	CE	48351.59	36.44
7	DF	50625.5	23.82
8	ES	49764.6	24.58
9	GO	53114.98	26.46
10	MA	31523.77	42.6
11	MG	270853.46	23.46
12	MS	19144.03	27.0
13	MT	29715.43	32.91

Key Insights:

- SP has the highest total freight value, which is expected as it has the largest customer base and the highest number of orders.
- SP also has the lowest average freight value, which is a positive sign. Lower shipping costs encourage customer retention and influence purchasing behaviour.
- More than seven states have an average freight value exceeding $40. Target should explore alternative shipping solutions to reduce costs, as high freight charges could drive customers to competitors.

Figure 11.1: Total and Average Freight Price by State



Sales, Freight, and Delivery Time Analysis

Question – 12: Find the number of days taken to deliver each order from the order's purchase date. Additionally, calculate the difference (in days) between the estimated and actual delivery dates of each order, using the following formulas:

- **time_to_deliver: order_delivered_customer_date - order_purchase_timestamp**
- **diff_estimated_delivery: order_delivered_customer_date - order_estimated_delivery_date**

To calculate the number of days taken to deliver each order and the difference between the actual and estimated delivery dates, I executed the following query:

```
SELECT
  order_id,
  DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp,
DAY) AS time_to_deliver,
  DATE_DIFF(order_delivered_customer_date,
order_estimated_delivery_date, DAY) AS diff_estimated_delivery
FROM `target.orders`
WHERE order_delivered_customer_date IS NOT NULL
```

This query calculates two metrics for each order:

- time_to_deliver: The number of days it took to deliver the order, calculated as the difference between the actual delivery date and the order purchase date.
- diff_estimated_delivery: The difference (in days) between the actual delivery date and the estimated delivery date.

Figure 12.1: Actual vs. Estimated Delivery Time

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	302bb8109d097a9fc6e9cefc5...	33	5
11	66057d37308e787052a32828...	38	6
12	19135c945c554eebfd7576c73...	36	2
13	4493e45e7ca1084efcd38ddeb...	34	0

Key Insights:

- The 'time_to_deliver' field shows the actual time taken for each order to be delivered, with values ranging from 29 to 67 days.
- The 'diff_estimated_delivery' field indicates whether the delivery was earlier or later than estimated.
- Negative values in the 'diff_estimated_delivery' column indicate orders delivered earlier than estimated, which could be seen as a positive for customer satisfaction.
- Longer-than-expected delivery times can indicate delays in the logistics process, which may affect customer satisfaction and require further investigation.

Question – 13: Identify the top five states with the highest and lowest average freight values.

To identify the top five states with the highest average freight values, I executed the following query:

```
SELECT
  DISTINCT(customer_state) AS state,
  ROUND(avg(freight_value),2) AS avg_freight_value,
  DENSE_RANK() OVER(ORDER BY avg(freight_value) DESC) AS top_tank
FROM `target.order_items` AS OI INNER JOIN `target.orders` AS O ON
OI.order_id = O.order_id
INNER JOIN `target.customers` AS C ON C.customer_id = O.customer_id
GROUP BY state
QUALIFY top_tank<=5
ORDER BY top_tank
```

This query calculates the average freight value for each state and ranks them based on the highest values. The top five states with the highest average freight costs are then selected.

Figure 13.1: Top Performing Five States by Average Freight Value (Query Output)

Row	state	avg_freight_value	top_tank
1	RR	42.98	1
2	PB	42.72	2
3	RO	41.07	3
4	AC	40.07	4
5	PI	39.15	5

Key Insights:

- RR (Roraima) has the highest average freight value at 42.98, suggesting high logistics costs in this state, possibly due to its geographical location and transportation challenges.
- PB (Paraíba) and RO (Rondônia) also have high freight values, which could be indicative of long-distance shipping or limited infrastructure.
- States with high freight values might experience customer dissatisfaction due to additional shipping costs, which could affect order volumes and customer retention.
- Target could explore strategies to optimize freight costs in these regions, such as leveraging alternative transportation methods, bulk shipping discounts, or local distribution hubs.

To identify the top five states with lowest average freight values, I executed the following query:

```
SELECT
  DISTINCT(customer_state) AS state,
  ROUND(avg(freight_value),2) AS avg_freight_value,
  DENSE_RANK() OVER(ORDER BY avg(freight_value)) AS bottom_rank
FROM `target.order_items` AS OI INNER JOIN `target.orders` AS O ON
OI.order_id = O.order_id
INNER JOIN `target.customers` AS C ON C.customer_id = O.customer_id
GROUP BY state
QUALIFY bottom_rank<=5
ORDER BY bottom_rank
```

This query calculates the average freight value for each state and ranks them based on the lowest values. The bottom five states with the lowest average freight costs are then selected.

Figure 13.2: Under Performing Five States by Average Freight Value (Query Output)

Row	state	avg_freight_value	bottom_rank
1	SP	15.15	1
2	PR	20.53	2
3	MG	20.63	3
4	RJ	20.96	4
5	DF	21.04	5

Figure 5.2.2

Key Insights:

- SP (São Paulo) has the lowest average freight value at 15.15, which could be attributed to its high shipping volume and well-established logistics infrastructure.
- States like PR (Paraná) and MG (Minas Gerais) also have relatively low freight costs, which could indicate proximity to major distribution centers or optimized delivery routes.
- Lower freight costs may contribute to higher customer satisfaction as customers are less likely to be deterred by expensive shipping charges.
- Target could leverage its strong performance in these regions to promote cost-effective delivery options in their marketing efforts, further boosting customer loyalty and order volumes.

Question – 14: Identify the top five states with the highest and lowest average delivery values.

To identify the top five states with the shortest average delivery times (best-performing states), I executed the following query:

```
SELECT
  DISTINCT(customer_state) AS state,
ROUND(AVG(TIMESTAMP_DIFF(order_delivered_customer_date,order_purchase
_timestamp,day)),2) AS avg_days,
  RANK() OVER(ORDER BY
AVG(TIMESTAMP_DIFF(order_delivered_customer_date,order_purchase_times
tamp,day))) AS actual_delivery_rank
FROM `target.orders` AS O INNER JOIN `target.customers` AS C ON
C.customer_id = O.customer_id
WHERE order_delivered_customer_date IS NOT NULL
GROUP BY state
QUALIFY actual_delivery_rank<=5
ORDER BY avg_days
```

This query calculates the average number of days between the order purchase date and the actual delivery date for each state. It then ranks the states based on delivery speed, with the top five states having the fastest deliveries.

Figure 14.1: Top Performing Five States by Average Delivery Days (Query Output)

Row	state	avg_days	actual_delivery_rank
1	SP	8.3	1
2	PR	11.53	2
3	MG	11.54	3
4	DF	12.51	4
5	SC	14.48	5

Key Insights:

- SP (São Paulo) has the fastest average delivery time at just 8.3 days, likely due to its well-developed logistics infrastructure and high order volume.
- The top-performing states generally have major urban centers, suggesting that proximity to distribution hubs plays a key role in delivery speed.
- Shorter delivery times lead to higher customer satisfaction and increased likelihood of repeat purchases. Target should leverage this advantage in these regions.

To identify the top five states with the longest average delivery times (where customers experience the most delays), I executed the following query:

```
SELECT
  DISTINCT(customer_state) AS state,
ROUND(AVG(TIMESTAMP_DIFF(order_delivered_customer_date,order_purchase
_timestamp,day)),2) AS avg_days,
  RANK() OVER(ORDER BY
AVG(TIMESTAMP_DIFF(order_delivered_customer_date,order_purchase_times
tamp,day)) DESC) AS actual_delivery_rank
FROM `target.orders` AS O INNER JOIN `target.customers` AS C ON
C.customer_id = O.customer_id
WHERE order_delivered_customer_date IS NOT NULL
GROUP BY state
QUALIFY actual_delivery_rank<=5
ORDER BY avg_days DESC
```

This query calculates the average number of days between the order purchase date and the actual delivery date for each state. It ranks the states based on delivery delays, with the top five states having the slowest deliveries.

Figure 14.2: Under Performing Five States by Average Delivery Days (Query Output)

Row	state	avg_days	actual_delivery_rank
1	RR	28.98	1
2	AP	26.73	2
3	AM	25.99	3
4	AL	24.04	4
5	PA	23.32	5

Key Insights:

- RR (Roraima) has the longest average delivery time at 28.98 days, indicating significant logistical challenges in this region.
- AP (Amapá) and AM (Amazonas) also have long delivery times, likely due to their remote locations and transportation difficulties.
- These states may require better logistics solutions, such as additional distribution centers, optimized carrier networks, or alternative delivery routes.
- Long delivery times could negatively impact customer satisfaction and retention. Target should explore strategies to improve delivery efficiency in these regions.

Question – 15: Identify the top five states where order delivery is significantly faster than the estimated delivery date, using the difference between the averages of actual and estimated delivery dates.

To identify the top five states where order delivery is significantly faster than the estimated delivery date, I executed the following query:

```
SELECT
  DISTINCT(customer_state) AS state,
ROUND(AVG(TIMESTAMP_DIFF(order_estimated_delivery_date,order_delivere
d_customer_date,day)),2) AS avg_days,
  RANK() OVER(ORDER BY
AVG(TIMESTAMP_DIFF(order_estimated_delivery_date,order_delivered_cust
omer_date,day)) DESC) AS order_delivery_rank
FROM `target.orders` AS O INNER JOIN `target.customers` AS C ON
C.customer_id = O.customer_id
WHERE order_delivered_customer_date IS NOT NULL
GROUP BY state
QUALIFY order_delivery_rank<=5
ORDER BY avg_days DESC
```

The query calculates the average difference (in days) between the estimated and actual delivery dates for each state. It then ranks the states based on how early deliveries occur, with the top five states selected for further analysis.

Figure 15.1: Top Five States with Significantly Faster Than Estimated Delivery

Row	state	avg_days	order_delivery_rank
1	AC	19.76	1
2	RO	19.13	2
3	AP	18.73	3
4	AM	18.61	4
5	RR	16.41	5

Key Insights:

- Orders in AC are delivered almost 20 days earlier than estimated, making it the state with the most significant early deliveries.
- All five states show a consistent pattern of deliveries arriving well before the estimated date, which could indicate efficient logistics, lower demand fluctuations, or overestimation in delivery projections.
- This trend presents an opportunity for Target to refine its estimated delivery date algorithm to improve accuracy and set better customer expectations.

Payment Analysis

Question – 16: Find the month-on-month number of orders placed using different payment types.

To analyst the month-on-month number of orders placed using different payment types, I utilized the following query:

```
WITH CTE AS (SELECT
  DISTINCT(payment_type) as payment_type,
  EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
  COUNT(DISTINCT(P.order_id)) as no_of_orders,
FROM `target.payments` AS P INNER JOIN `target.orders` AS O ON
P.order_id = O.order_id
GROUP BY payment_type, month
ORDER BY payment_type,month)

SELECT
  payment_type,
  month,
  no_of_orders,
  LAG(no_of_orders,1) OVER(PARTITION BY payment_type ORDER BY month)
previous_month_orders
FROM CTE
```

The query first extracts the month and payment type and counts the number of unique orders for each category. Using a Common Table Expression (CTE), it then calculates the number of orders from the previous month for comparison.

Figure 16.1: month-on-month number of orders placed using different payment types.

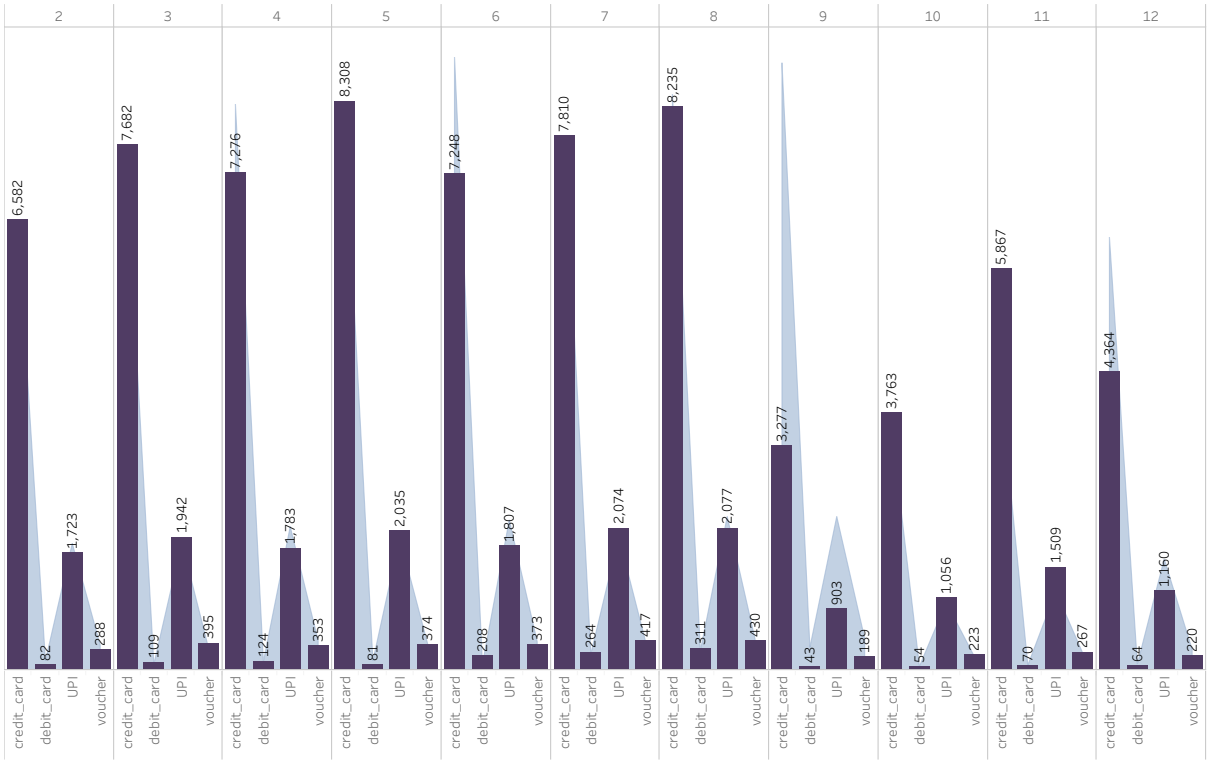
Row	payment_type	month	no_of_orders	previous_month_ord
1	not_defined	8	2	null
2	not_defined	9	1	2
3	debit_card	1	118	null
4	debit_card	2	82	118
5	debit_card	3	109	82
6	debit_card	4	124	109
7	debit_card	5	81	124
8	debit_card	6	208	81
9	debit_card	7	264	208
10	debit_card	8	311	264
11	debit_card	9	43	311
12	debit_card	10	54	43
13	debit_card	11	70	54

Key Insights:

- Credit card payments are the most preferred payment method, with a consistently high number of orders across multiple months.

- UPI payments show a steady month-on-month increase, indicating growing adoption of digital payments. This trend suggests that Target could benefit from offering incentives for UPI transactions.

Figure 16.2: Order Payment Type Analysis (MoM Comparison)



Note: The bar chart indicates current month's sales, while the area chart indicates previous month's sales.

Question – 17: Determine the number of orders placed based on the payment instalments made.

To determine the number of orders placed based on the payment instalments made, I utilised the following query:

```
SELECT
    payment_installments,
    COUNT(DISTINCT order_id) AS num_orders
FROM
    target.payments
GROUP BY
    payment_installments
ORDER BY
    installments_used;
```

The query retrieves the number of installments used and the corresponding number of orders from the payments table.

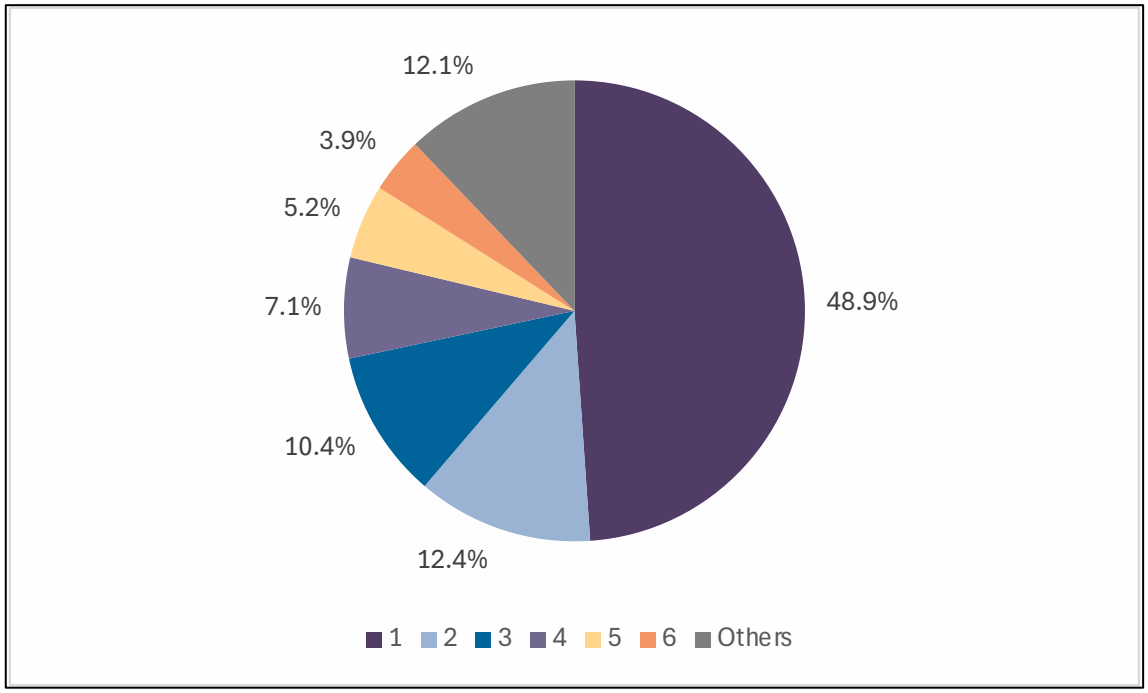
Figure 17.1: Distribution of Orders Across Installment Plans (Query Output)

Row	payment_installment	num_orders
1	0	2
2	1	49060
3	2	12389
4	3	10443
5	4	7088
6	5	5234
7	6	3916
8	7	1623
9	8	4253
10	9	644
11	10	5315
12	11	23
13	12	133

Key Insights:

- A single installment is the most preferred payment option, indicating that most customers opt for a one-time cash or card payment. This is a positive sign for Target, as it suggests financial stability among customers and lower risks associated with installment-based payments.
- Nearly 99.6% of customers make payments in 10 installments, highlighting a strong preference for long-term financing. This trend could be leveraged to offer installment-based promotions or optimize financing options.

Figure 17.2: Percentage of Orders by Installment Plan



Note: The numbers indicate the number of installments customers selected

References

1. Devaraj, V. (2024) Target Dataset. Available at: https://www.kaggle.com/datasets/devarajv88/target-dataset (Accessed: 21 February 2025).
2. PR Newswire. (2024) Target Corporation Reports Fourth Quarter and Full Year 2023 Earnings. Available at: https://www.prnewswire.com/news-releases/target-corporation-reports-fourth-quarter-and-full-year-2023-earnings-302079218.html#:~:text=Target%20Corporation%20(NYSE%3A%20TGT),compared%20with%20%241.89%20in%202022 (Accessed: 21 February 2025).
3. Scaler. (2024) Business Case: Target SQL (Accessed: 23 December 2024)
4. Target Corporation. (2024) 2023 Annual Report. Available at: https://corporate.target.com/investors/annual/2023-annual-report (Accessed: 21 February 2025).