TARGET SQL CASE STUDY PROJECT

Target Corporation is a major American retail company, known for its chain of discount department stores and hypermarkets. It's the seventh-largest retailer in the United States. Target is headquartered in Minneapolis, Minnesota, and operates nearly 2,000 stores across all fifty states and the District of Columbia. Target evolved from the Dayton Dry Goods Company, which began in 1902. It transitioned to discount retailing in the 1960s and eventually became the largest division of Dayton Hudson Corporation, which later changed its name to Target Corporation in 2000. This document is a case study on Target is concerned with the data of their outlet in Brazil. It will depict actionable insights and help in understanding various strategies that can be implemented for growth and expansion of the business. It can also shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.
The dataset consists of 8 different tables containing data in a sorted manner for better understanding. The tables are as follows:

1. Customers table - Contains customers' general information.
2. Sellers table - Contains data of the sellers who provide goods to the outlet.
3. Products table - Contains product data that are bought and been sold.
4. Orders table - Contains data of orders placed by the customers.
5. Payments table - Contains data of all the payments made by the customers on purchasing.
6. Geolocation - The geolocation table maps unique postal codes (geolocation_zip_code_prefix) to their corresponding latitude, longitude, city, and state, enabling spatial analysis of customer and seller locations across Brazil.
7. Order Items table - Contains data of the products that are being ordered by customers including the shipping date and price of the product.
8. Order Reviews table - Contains data on reviews given by the customers who had placed orders.

- ## **DATA DEFINITION:**

CUSTOMERS TABLE:

The customers table consists of 99441 rows and 5 columns. (customer_id, customer_unique_id, customer_zip_code_prefix, customer_city, customer_state)

```
SELECT *
FROM `scaler-dsml-sql-465408.target.customers`
```

Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	customer_id	customer_unique_id	customer_zip_co...	customer_city	customer_state
1	2201362e68992f654942dc006...	f7d7fc0a59ef4363fdce6e3aa06...	69900	rio branco	AC
2	31dbc13addc753e210692eaca...	5dbba6c01268a8ad43f79157bf...	69900	rio branco	AC
3	dad907e170748a35ef4e92238...	36b1c0516f123351ffa8743041...	69900	rio branco	AC
4	888d2ebe1af2a8c93c75dae5df...	721d1092e1a6460c67e6a0e69...	69900	rio branco	AC
5	8a0108267d9258a0ec9f74381...	7a2dc4682890550ebe3b8befce...	69900	rio branco	AC
6	5880e46677c68394bda62479f...	5dbba6c01268a8ad43f79157bf...	69900	rio branco	AC
7	53996870173a3a001a1fb56ef0...	2624230437101e0bdcea1e483...	69900	rio branco	AC
8	cd281c1a7d26cd29a3ed4b029f...	086d6b5b5ba195a91aa0a6ec8...	69900	rio branco	AC
9	717a3a9459f9006f23e24a684b...	5203034a29c3d294713bf8ccb...	69900	rio branco	AC
10	2d618e470c95c9b425cbb0cbc...	f7d7fc0a59ef4363fdce6e3aa06...	69900	rio branco	AC

SELLERS TABLE:

The sellers table consists of 3095 rows and 4 columns. (seller_id ,seller_zip_code_prefix, seller_city ,seller_state)

```
SELECT *
FROM `scaler-dsml-sql-465408.target.sellers`
```

Query results

Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	seller_id	seller_zip_code_p...	seller_city	seller_state
1	c13ef0cfbe42f190780f621ce81...	1207	sao paulo sp	SP
2	5444b12c82f21c923f2639ebc7...	2051	sao pauo	SP
3	1cbd32d00d01bb8087a5eb088...	3363	sp / sp	SP
4	71593c7413973a1e160057b80...	3407	sao paulo / sao paulo	SP
5	6f1a1263039c76e68f40a8e536...	3581	sao paulop	SP
6	06579cb253ecd5a3a12a9e6eb...	4007	sao paulo - sp	SP
7	8090490573c6c0aa343a7231e...	4130	sao paulo - sp	SP
8	a3fa18b3f688ec0fca3eb8bfcbd...	4557	são paulo	SP
9	2156f2671501a81034d7d07f21...	4776	sp	SP
10	0b18d63d0cd1d723567903fd3...	5141	sp	SP

PRODUCTS TABLE:

The products table consists of 32951 rows and 9 columns.(product_id, product_category, product_name_length, product_description_length, product_photos_qty, product_weight_g, product_length_cm, product_height_cm, product_width_cm)

```sql
SELECT *
FROM `scaler-dsml-sql-465408.target.products`
```

Query results

Save results ▾ Open in ▾

Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	product_id	product category	product_name_length	product_description_length	product_photos_qty	product_weight_g	product_length_cm	product_height_cm	product_width_cm
1	a0ab96e461d74537772b84950...	climatization	41	717	1	1050	18	7	8
2	20ae7c024ede613f47e0d2f23f...	fixed telephony	25	455	1	330	17	11	9
3	4d7585daba2f8b3ed7f8744790...	fixed telephony	53	897	2	300	15	8	9
4	ad7aebed205805125489f8a89...	Construction Tools Tools	41	2526	2	1150	22	10	9
5	980ecbcc15fe174ec1e5757c4d...	Agro Industria e Comercio	48	157	1	250	17	3	10
6	2b6535d32c6996c9478c131a8...	CONSTRUCTION SECURI...	50	428	2	333	16	9	10
7	16d096faa27582985f849f0837...	Agro Industria e Comercio	50	1153	1	4050	11	18	10
8	c8eca123751676bbdaaa9e4a8...	Construction Tools Tools	53	46	2	1520	22	20	11
9	8f4e5f5e6323576768b835363a...	Fashion Men's Clothing	29	50	1	300	30	40	11
10	170ae15fc78b44bceab3e66bae...	Fashion Men's Clothing	38	50	1	300	30	40	11

ORDERS TABLE:

The orders table consists of 99441 rows and 8 columns.(order_id, customer_id, order_status, order_purchase_timestamp, order_approved_at, order_delivered_carrier_date, order_delivered_customer_date, order_estimated_delivery_date)

```sql
SELECT *
FROM `scaler-dsml-sql-465408.target.orders`
```

Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	order_id	customer_id	order_stat...	order_purchase_timesta...	order_approved_at	order_delivered_carrier_...	order_delivered_custom...	order_estimated_delivery...
1	a2e4c44360b4a57bdff22f3a46...	8886130db0ea6e9e70ba0b03d...	approved	2017-02-06 20:18:17 UTC	2017-02-06 20:30:19 UTC	null	null	2017-03-01 00:00:00 UTC
2	132f1e724165a07f6362532bfb...	b2191912d8ad6eac2e4dc3b6e...	approved	2017-04-25 01:25:34 UTC	2017-04-30 20:32:41 UTC	null	null	2017-05-22 00:00:00 UTC
3	809a282bbd5dbcabb6f2f724fc...	622e13439d6b5a0b486c43561...	canceled	2016-09-13 15:24:19 UTC	2016-10-07 13:16:46 UTC	null	null	2016-09-30 00:00:00 UTC
4	e5215415bb6f76fe3b7cb68103...	b6f6cbfc126f1ae6723fe2f9b37...	canceled	2016-10-22 08:25:27 UTC	null	null	null	2016-10-24 00:00:00 UTC
5	71303d7e93b399f5bcd537d12...	b106b360fe2ef8849fbbd056f7...	canceled	2016-10-02 22:07:52 UTC	2016-10-06 15:50:56 UTC	null	null	2016-10-25 00:00:00 UTC
6	e5fa5a7210941f7d56d0208e4e...	683c54fc24d40ee9f8a6fc179fd...	canceled	2016-09-05 00:15:34 UTC	2016-10-07 13:17:15 UTC	null	null	2016-10-28 00:00:00 UTC
7	92d731517f17c26f16182d4549...	95c44daefa7bfb91c6bb0665a0...	canceled	2016-10-05 11:23:13 UTC	null	null	null	2016-11-14 00:00:00 UTC
8	ddaec6fff982b13e7e048b627a...	68f4ad79cc0c2ad06e19088f5c...	canceled	2016-10-04 19:41:32 UTC	null	null	null	2016-11-16 00:00:00 UTC
9	c549f0d88f33bfe43351b893f1f...	ea26950075411d4c3569a7185...	canceled	2016-10-06 13:21:05 UTC	null	null	null	2016-11-16 00:00:00 UTC
10	9bd5312d12f48b04a460f702ba...	4d145fd1d3245cbb3697dd41b...	canceled	2016-10-06 20:43:30 UTC	null	null	null	2016-11-16 00:00:00 UTC

PAYMENTS TABLE:

The payments table consists of 103886 rows and 5 columns. (order_id, payment_sequential, payment_type, payment_installments, payment_value)

```sql
SELECT *
FROM `scaler-dsml-sql-465408.target.payments`
```

Query results

Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	order_id	payment_sequential	payment_type	payment_installments	payment_value
1	744bade1fcf9ff3f31d860ace07...	2	credit_card	0	58.69
2	1a57108394169c0b47d8f876a...	2	credit_card	0	129.94
3	8bcbe01d44d147f901cd31926...	4	voucher	1	0.0
4	fa65dad1b0e818e3ccc5cb0e39...	14	voucher	1	0.0
5	6ccb433e00daae1283ccc9561...	4	voucher	1	0.0
6	4637ca194b6387e2d538dc89b...	1	not_defined	1	0.0
7	00b1cb0320190ca0daa2c88b3...	1	not_defined	1	0.0
8	45ed6e85398a87c253db47c2d...	3	voucher	1	0.0
9	fa65dad1b0e818e3ccc5cb0e39...	13	voucher	1	0.0
10	c8c528189310eaa44a745b8d9...	1	not_defined	1	0.0

GEOLOCATION:

The geolocation table consists of 1000163 rows and 5 columns. (geolocation_zip_code_prefix, geolocation_lat, geolocation_lng, geolocation_city, geolocation_state)

```sql
SELECT *
FROM `scaler-dsml-sql-465408.target.geolocation`
```

Query results

Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	geolocation_zip_...	geolocation_lat	geolocation_lng	geolocation_city	geolocation_state
1	1037	-23.5456212811...	-46.6392920480...	sao paulo	SP
2	1046	-23.5460811270...	-46.6448202983...	sao paulo	SP
3	1046	-23.5461289664...	-46.6429514836...	sao paulo	SP
4	1041	-23.5443921648...	-46.6394993062...	sao paulo	SP
5	1035	-23.5415779617...	-46.6416072232...	sao paulo	SP
6	1012	-23.5477623033...	-46.6353605378...	são paulo	SP
7	1029	-23.5437690557...	-46.6342778408...	sao paulo	SP
8	1011	-23.5476395503...	-46.6360316231...	sao paulo	SP
9	1014	-23.5464353433...	-46.6338302339...	sao paulo	SP
10	1012	-23.5489459851...	-46.6346711329...	sao paulo	SP

ORDER ITEMS TABLE:

The order items table consists of 112650 rows and 7 columns. (order_id, order_item_id, product_id, seller_id, shipping_limit_date, price, freight_value)

```sql
SELECT *
FROM `scaler-dsml-sql-465408.target.order_items`
```

Query results

Save results

Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	order_id	order_item_id	product_id	seller_id	shipping_limit_date	price	freight_value
1	3ee6513ae7ea23bdfab5b9ab60...	1	8a3254bee785a526d548a81a9...	96804ea39d96eb908e7c3afdb6...	2018-05-04 03:55:26 UTC	0.85	18.23
2	6e864b3f0ec71031117ad4cf46...	1	8a3254bee785a526d548a81a9...	96804ea39d96eb908e7c3afdb6...	2018-05-02 20:30:34 UTC	0.85	18.23
3	c5bdd8ef3c0ec420232e668302...	2	8a3254bee785a526d548a81a9...	96804ea39d96eb908e7c3afdb6...	2018-05-07 02:55:22 UTC	0.85	22.3
4	8272b63d03f5f79c56e9e4120a...	2	05b515fdc76e888aada3c6d66c...	2709af9587499e95e803a6498...	2017-07-21 18:25:23 UTC	1.2	7.89
5	8272b63d03f5f79c56e9e4120a...	3	05b515fdc76e888aada3c6d66c...	2709af9587499e95e803a6498...	2017-07-21 18:25:23 UTC	1.2	7.89
6	8272b63d03f5f79c56e9e4120a...	4	05b515fdc76e888aada3c6d66c...	2709af9587499e95e803a6498...	2017-07-21 18:25:23 UTC	1.2	7.89
7	8272b63d03f5f79c56e9e4120a...	5	05b515fdc76e888aada3c6d66c...	2709af9587499e95e803a6498...	2017-07-21 18:25:23 UTC	1.2	7.89
8	8272b63d03f5f79c56e9e4120a...	6	05b515fdc76e888aada3c6d66c...	2709af9587499e95e803a6498...	2017-07-21 18:25:23 UTC	1.2	7.89
9	8272b63d03f5f79c56e9e4120a...	7	05b515fdc76e888aada3c6d66c...	2709af9587499e95e803a6498...	2017-07-21 18:25:23 UTC	1.2	7.89
10	8272b63d03f5f79c56e9e4120a...	8	05b515fdc76e888aada3c6d66c...	2709af9587499e95e803a6498...	2017-07-21 18:25:23 UTC	1.2	7.89

ORDER REVIEWS TABLE:

The order reviews table consists of 99224 rows and 6 columns. (review_id, order_id, review_score, review_comment_title, review_creation_date, review_answer_timestamp)

```sql
SELECT *
FROM `scaler-dsml-sql-465408.target.order_reviews`
```

Query results

Save results

Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	review_id	order_id	review_score	review_comment_title	review_creation_date	review_answer_timestamp
1	9e628acdf082a5eeea001b1ded...	54d744a4410b1edccc36c6d1f1...	1	null	0001-02-17 00:00:00 UTC	0015-03-17 02:47:00 UTC
2	db27636ea82d449a9e0d1bea6...	a9a93c428c6103f2151bb63a1...	1	null	0001-02-17 00:00:00 UTC	0001-02-17 11:12:00 UTC
3	5bec0620e043423a063641e9a...	1cb796218c383fc54a6a45414...	1	null	0001-02-17 00:00:00 UTC	0013-02-17 11:43:00 UTC
4	3c9d8bbd2a92baddc2356f4c8e...	b3feb3846bb0a8d68cd328138...	1	null	0001-02-18 00:00:00 UTC	0002-02-18 10:36:00 UTC
5	647a5c7d693aedb5ce12f3c5e5...	745e2506fb647deca4669e1c88...	1	null	0001-02-18 00:00:00 UTC	0001-02-18 21:10:00 UTC
6	eff3aab917323a17b7a71daeb5...	ccd6dca45b65f311ff32ed8fbfd...	1	null	0001-02-18 00:00:00 UTC	0001-02-18 18:50:00 UTC
7	c3a3e991eaf0805592aa797a14...	057bc88d7845bc736fbe1fb27f...	1	null	0001-02-18 00:00:00 UTC	0002-02-18 12:34:00 UTC
8	8f28504ad064a22758b7ae037...	9bc5c73ed95bec088ff04d9249...	1	null	0001-02-18 00:00:00 UTC	0006-02-18 21:50:00 UTC
9	061a08afa5057258bc7aea4d6...	748d3b50f6f4b32e9998bb6282...	1	null	0001-02-18 00:00:00 UTC	0002-02-18 06:55:00 UTC
10	08bdff34c7de564b6fb2312252...	61dffbf220e7df65d65f1ea067f...	1	null	0001-02-18 00:00:00 UTC	0001-02-18 23:04:00 UTC

DATA TYPES OF ALL COLUMNS IN CUSTOMERS TABLE:

```sql
SELECT
  column_name,data_type
FROM `scaler-dsml-sql-465408.target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'customers';
```

Job information | Results | Visualization | JSON | Execution details

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

DATA TYPES OF ALL COLUMNS IN PRODUCTS TABLE:

```sql
SELECT
  column_name,data_type
FROM `scaler-dsml-sql-465408.target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'products';
```

Job information | Results | Visualization | JSON | E

Row	column_name	data_type
1	product_id	STRING
2	product category	STRING
3	product_name_length	INT64
4	product_description_length	INT64
5	product_photos_qty	INT64
6	product_weight_g	INT64
7	product_length_cm	INT64
8	product_height_cm	INT64
9	product_width_cm	INT64

DATA TYPE OF ALL COLUMNS IN SELLERS TABLE:

```sql
SELECT
  column_name,data_type
FROM `scaler-dsml-sql-465408.target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'sellers';
```

Job information | Results | Visualization | JSON

Row	column_name	data_type
1	seller_id	STRING
2	seller_zip_code_prefix	INT64
3	seller_city	STRING
4	seller_state	STRING

DATA TYPES OF ALL COLUMNS IN ORDERS TABLE:

```sql
SELECT
  column_name,
  data_type
FROM `scaler-dsml-sql-465408.target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'orders';
```

Job information | Results | Visualization | JSON | E

Row	column_name	data_type
1	order_id	STRING
2	customer_id	STRING
3	order_status	STRING
4	order_purchase_timestamp	TIMESTAMP
5	order_approved_at	TIMESTAMP
6	order_delivered_carrier_date	TIMESTAMP
7	order_delivered_customer_date	TIMESTAMP
8	order_estimated_delivery_date	TIMESTAMP

DATA TYPES OF ALL COLUMNS IN PAYMENTS TABLE:

```sql
SELECT
  column_name,
  data_type
FROM `scaler-dsml-sql-465408.target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'orders';
```

Row	column_name	data_type
1	order_id	STRING
2	payment_sequential	INT64
3	payment_type	STRING
4	payment_installments	INT64
5	payment_value	FLOAT64

DATA TYPES OF ALL COLUMNS IN GEOLOCATION TABLE:

```sql
SELECT
  column_name,
  data_type
FROM `scaler-dsml-sql-465408.target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'geolocation';
```

Job information | Results | Visualization | JSON

Row	column_name	data_type
1	geolocation_zip_code_prefix	INT64
2	geolocation_lat	FLOAT64
3	geolocation_lng	FLOAT64
4	geolocation_city	STRING
5	geolocation_state	STRING

DATA TYPES OF ALL COLUMNS IN ORDER ITEMS TABLE:

```
SELECT
  column_name,
  data_type
FROM `scaler-dsml-sql-465408.target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'order_items';
```

Job information Results Visualization JSON

Row	column_name	data_type
1	order_id	STRING
2	order_item_id	INT64
3	product_id	STRING
4	seller_id	STRING
5	shipping_limit_date	TIMESTAMP
6	price	FLOAT64
7	freight_value	FLOAT64

DATA TYPES OF ALL COLUMNS IN ORDER REVIEWS TABLE:

```
SELECT
  column_name,
  data_type
FROM `scaler-dsml-sql-465408.target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'order_reviews';
```

Job information Results Visualization JSON E

Row	column_name	data_type
1	review_id	STRING
2	order_id	STRING
3	review_score	INT64
4	review_comment_title	STRING
5	review_creation_date	TIMESTAMP
6	review_answer_timestamp	TIMESTAMP

- **<u>DATA ANALYSIS:</u>**

1. <u>TOP 10 STATES IN BRAZIL WITH MOST CUSTOMERS:</u>

The below query and table depicts the top 10 states in Brazil with the most number of customers.

QUERY:

```
SELECT
  customer_state,
  COUNT(DISTINCT customer_unique_id) AS num_customers
FROM `scaler-dsml-sql-465408.target.customers`
GROUP BY customer_state
ORDER BY num_customers DESC
LIMIT 10;
```

RESULT:

Job information | Results | Visualization

Row	customer_state	num_customers
1	SP	40302
2	RJ	12384
3	MG	11259
4	RS	5277
5	PR	4882
6	SC	3534
7	BA	3277
8	DF	2075
9	ES	1964
10	GO	1952

INSIGHTS

- ➢ The top states - very likely São Paulo (SP), Rio de Janeiro (RJ), Minas Gerais (MG), Rio Grande do Sul (RS, Paraná (PR)account for a disproportionately high share of Target's Brazilian customers.This shows a regional concentration. In short, urban/metropolitan areas dominate in e-commerce penetration.
- ➢ Conversely, states that don't appear in the top 10 have untapped potential — smaller but growing digital adoption.

RECOMMENDATIONS

- ➢ Set up or expand fulfillment centers and local delivery hubs in the top customer states to reduce delivery time and freight cost.
- ➢ Run regional promotions (free shipping, discounts) in emerging states to boost adoption.
- ➢ Partner with local sellers in smaller states to reduce freight costs and increase product relevance.

2. <u>TIME RANGE BETWEEN ORDERS PLACED:</u>

The below query and table depicts the time range between the first and last date of orders placed in Brazil.

QUERY:

```sql
SELECT
MIN(order_purchase_timestamp) AS first_order_date,
MAX(order_purchase_timestamp) AS last_order_date
FROM `scaler-dsml-sql-465408.target.orders`;
```

RESULTS:

Query results

Job information | Results | Visualization | JSON

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

INSIGHTS

- ➢ The first order was placed on 4th of September 2016 and last order on 17th of October 2018. The data covers a timespan of 2 whole years which is helpful enough for further analysis.
- ➢ The timeframe includes key retail events like Christmas, Black Friday (Nov), and Mother's Day (May). These are likely to show order spikes and can reveal seasonal shopping behavior.

RECOMMENDATIONS

- ➢ Target can use 2017 as the baseline year to compare 2018 (Jan–Oct) with 2017 to measure growth in orders, sales value, and customer base.
- ➢ Since the data allows tracking customers over multiple years, Target should analyze repetitive vs new customers to design loyalty campaigns.
- ➢ Target should collect latest order data (2019 onwards) for more accurate decision-making, since the data contains information till Oct 2018.

3. <u>COUNT OF UNIQUE CITIES AND STATES FROM WHERE ORDERS WERE PLACED:</u>

The below query and table helps to understand the total number of unique cities and states from where the customers have placed their orders within the 2 years of data that we have.

QUERY:

```
SELECT
COUNT(DISTINCT c.customer_city) AS unique_cities,
COUNT(DISTINCT c.customer_state) AS unique_states
FROM `scaler-dsml-sql-465408.target.customers` AS c
JOIN `scaler-dsml-sql-465408.target.orders` AS o
USING(customer_id)
WHERE order_purchase_timestamp IS NOT NULL;
```

RESULTS:

Query results

Job information | Results | Visualization

Row	unique_cities	unique_states
1	4119	27

INSIGHTS

- ➢ Target's e-commerce presence spans over 27 states and over 4,100 cities in Brazil, indicating a strong national reach.
- ➢ The very large number of unique cities (4119) compared to only 27 states suggests that customers are not just concentrated in big urban hubs, but also spread across smaller towns.

RECOMMENDATIONS

- ➢ Target should invest in regional warehouses in states with high customer density to reduce freight cost & delivery time.
- ➢ Target can identify the top 10–20 cities by order volume and customer base and consider offering same-day / next-day delivery in those major hubs to improve customer satisfaction.
- ➢ Also, states with fewer customers could be tapped with targeted promotions and free shipping offers to grow market penetration.

4. <u>COUNT OF CUSTOMERS PLACING ORDERS FROM EACH CITY AND STATE:</u>

The below query and table helps to discover the unique number of customers belonging to each city and state in order to understand the popularity/reach of Target within each part of the country. We get 4310 rows of output for this query.

QUERY:

```
SELECT
c.customer_state,
c.customer_city,
COUNT(DISTINCT c.customer_unique_id) AS customer_count
FROM `scaler-dsml-sql-465408.target.orders` AS o
JOIN `scaler-dsml-sql-465408.target.customers` AS c
USING(customer_id)
WHERE o.order_purchase_timestamp IS NOT NULL
GROUP BY c.customer_state,c.customer_city
ORDER BY customer_count DESC;
```

RESULTS:

Job information Results Visualization JSON Execution details

Row	customer_state	customer_city	customer_count
1	SP	sao paulo	14984
2	RJ	rio de janeiro	6620
3	MG	belo horizonte	2672
4	DF	brasilia	2069
5	PR	curitiba	1465
6	SP	campinas	1398
7	RS	porto alegre	1326
8	BA	salvador	1209
9	SP	guarulhos	1153
10	SP	sao bernardo do campo	908

INSIGHTS

- ➢ A handful of big metropolitan cities (like São Paulo, Rio de Janeiro, Belo Horizonte) dominate the customer count.
- ➢ Smaller cities contribute less which means that revenue and customer base are highly dependent on urban hubs. Cities with high unique customer counts also likely drive repeat orders, since urban customers tend to purchase more frequently online.

RECOMMENDATIONS

- ➢ Target should prioritize São Paulo, Rio de Janeiro, Belo Horizonte, Brasilia for targeted campaigns, faster delivery promises, and exclusive offers. These hubs provide the highest ROI for ad spend due to dense customer base.
- ➢ It can promote fashion & electronics in urban areas, while pushing home essentials & low-ticket items in smaller cities.

5. <u>STATE-WISE RANKING OF UNIQUE CUSTOMERS ALONG WITH TOTAL ORDERS PLACED:</u>

The breakdown of total number of unique customers along with the total orders placed within each state is depicted below with the help of a query and table below.

QUERY:

```
SELECT
  c.customer_state,
  COUNT(DISTINCT c.customer_unique_id) AS unique_customers,
  COUNT(o.order_id) AS total_orders
FROM `scaler-dsml-sql-465408.target.customers` AS c
JOIN `scaler-dsml-sql-465408.target.orders` AS o
USING(customer_id)
WHERE o.order_purchase_timestamp IS NOT NULL
GROUP BY c.customer_state
ORDER BY unique_customers DESC;
```

RESULTS:

Job information Results Visualization JSON Execution c

Row	customer_state	unique_customers	total_orders
1	SP	40302	41746
2	RJ	12384	12852
3	MG	11259	11635
4	RS	5277	5466
5	PR	4882	5045
6	SC	3534	3637
7	BA	3277	3380
8	DF	2075	2140
9	ES	1964	2033
10	GO	1952	2020

INSIGHTS

- ➢ States like São Paulo (SP), Rio de Janeiro (RJ), Minas Gerais (MG), and Rio Grande do Sul (RS) are dominating both unique customers and total orders. This shows Target's core market is concentrated in a few high-population, high-income states.
- ➢ Northern and rural states (e.g., Amazonas, Acre, Roraima) likely have lower unique customers. This indicates a geographic penetration gap either due to weak logistics, lower trust in online shopping, or less marketing focus.

RECOMMENDATIONS

- ➢ For top states like SP, RJ and MG, Target should push loyalty programs, subscription models, and faster delivery guarantees.
- ➢ They can also collaborate with regional courier companies to reduce delivery challenges in low-volume states. Faster deliveries will help build trust and repeat purchases outside urban hubs.
- ➢ For lower-ranked states, Target should analyze logistics functionality and offer introductory free shipping or regionalized discounts to build trust in first-time buyers.

6. <u>GROWING TREND IN THE NUMBER OF ORDERS PLACED OVER PAST YEARS:</u>

The trend of orders placed within each month of the year in 2 years is demonstrated below with the help of an SQL query and a table. This table depicts the first 10 rows of output out of the total 25 rows.

QUERY:

```
SELECT
EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
EXTRACT(MONTH FROM order_purchase_timestamp) AS order_month,
COUNT(order_id) AS total_orders
FROM `scaler-dsml-sql-465408.target.orders`
WHERE order_purchase_timestamp IS NOT NULL
GROUP BY order_year, order_month
ORDER BY order_year, order_month;
```

RESULTS:

Job information | Results | Visualization | JSON

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

INSIGHTS

- ➢ Orders have grown significantly from late 2016 through 2017 and into 2018. This shows Target is gaining traction with customers and expanding reach over time.
- ➢ While comparing the same months across years reveals sustained growth. (e.g., Nov 2017 > Nov 2016, and Nov 2018 > Nov 2017)

RECOMMENDATIONS

- ➢ Target can run discount campaigns, loyalty rewards, or free shipping offers in low months to smoothen demand cycles.
- ➢ They can double down on marketing and promotions during high-demand months (holidays, festivals). Partner with sellers to launch seasonal offers and ensure inventory preparedness.
- ➢ Target can see which regions drive seasonal growth and target those areas with localized promotions.

7. <u>MONTHLY SEASONALITY IN TERMS OF ORDERS BEING PLACED:</u>

The month-wise seasonality of orders placed by customers throughout is presented below with the help of an SQL query and a table. The output consists of 2 years of month-wise data.

QUERY:

```
SELECT
EXTRACT(MONTH FROM order_purchase_timestamp) AS order_month,
COUNT(order_id) AS total_orders
FROM `scaler-dsml-sql-465408.target.orders`
WHERE order_purchase_timestamp IS NOT NULL
GROUP BY order_month
ORDER BY order_month;
```

RESULTS:

Row	order_month	total_orders
1	1	8069
2	2	8508
3	3	9893
4	4	9343
5	5	10573
6	6	9412
7	7	10318
8	8	10843
9	9	4305
10	10	4959
11	11	7544
12	12	5674

INSIGHTS

➢ May (10,573), July (10,318), August (10,843) show the highest order volumes, indicating that mid-year months drive the most sales likely due to festivals or seasonal demand.

➢ Orders are lower in January (8,069), February (8,508), and April (9,343).

➢ September (4,305) and October (4,959) are significantly lower than other months.

➢ November (7,544) and December (5,674) are not the biggest sales months. This suggests that Brazilian customers don't rely heavily on end-of-year shopping.

RECOMMENDATIONS

➢ Target can introduce seasonal campaigns or flash sales in Sep–Oct to smooth out the demand gap. Eg: "Back-to-School Sales" Or "Pre-Holiday Shopping Discounts."

➢ Since mid-year is naturally strong, Target can double down with high-visibility campaigns, seller discounts, and aggressive marketing on popular categories (electronics, fashion, home appliances).

➢ Also for Jan–Feb and Sep–Oct, Target can focus on loyalty programs, cashback rewards, or subscription models to retain engagement year-round.

8. <u>TIME AT WHICH BRAZILIAN CUSTOMERS HAVE MOSTLY PLACED ORDERS:</u>

The time (in hours) during which customers have mostly placed their orders is categorized into 4 intervals (0-6 hrs : Dawn, 7-12 hrs : Morning, 13-18 hrs : Afternoon, 19-23 hrs : Night). This is represented below with the help of a table and an SQL query.

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 'Mornings'
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 13 AND 18 THEN 'Afternoon'
ELSE 'Night'
END AS time_of_day,
COUNT(order_id) AS total_orders
FROM `scaler-dsml-sql-465408.target.orders`
WHERE order_purchase_timestamp IS NOT NULL
GROUP BY time_of_day
ORDER BY total_orders desc;
```

RESULTS:

Row	time_of_day	total_orders
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

INSIGHTS

- ➢ Afternoon & Night tend to dominate maybe due to customers shopping after work hours.
- ➢ Morning orders are moderate which is possibly workplace shopping.
- ➢ Dawn orders are the lowest which means that very few people shop between midnight to 6am.

RECOMMENDATIONS

- ➢ To stimulate sales throughout the day, Target can run promotions in Afternoon & Night to push discounts during peak browsing times. Improve app/site performance at night to ensure checkout flow is smooth when demand peaks.
- ➢ Use push notifications in Morning & Afternoon to capture early shoppers before peak hours.
- ➢ Test flash sales in low-demand Dawn hours to attract niche segments (night workers, insomniacs).

9. <u>MONTH-WISE ORDERS PLACED BY CUSTOMERS IN EACH STATE:</u>

The bifurcation of orders placed by customers from each state within Brazil in the past 2 years is represented below with the help of a table and an SQL query.

QUERY:

```
SELECT
c.customer_state,
EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
COUNT(o.order_id) AS total_orders
FROM `scaler-dsml-sql-465408.target.orders` AS o
JOIN `scaler-dsml-sql-465408.target.customers` AS c
USING(customer_id)
GROUP BY customer_state,order_month
ORDER BY customer_state,order_month;
```

RESULTS:

Row	customer_state	order_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

INSIGHTS

- ➢ Large states like São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG) likely dominate order volumes each month.
- ➢ Certain states like Paraná, Rio Grande do Sul may have steady month-on-month increases, showing adoption beyond the core urban centers.

RECOMMENDATIONS

- ➢ Target should invest more in São Paulo, RJ, and MG since they consistently generate the highest volume of sales.
- ➢ For growing smaller states, use localized campaigns (regional holidays, cultural festivals).
- ➢ Target should position warehouses closer to high-order states (SP, RJ, MG) to reduce delivery times.
- ➢ In states with stable but not growing demand, prioritize loyalty programs.

10. UNIQUE CUSTOMERS PRESENT IN EACH STATE:

The unique number of customers statewide in Brazil is represented below with the help of a table and an SQL query.

QUERY:

```
SELECT
customer_state,
COUNT(DISTINCT customer_id) AS customer_count
FROM `scaler-dsml-sql-465408.target.customers`
GROUP BY customer_state
ORDER BY customer_count DESC;
```

RESULTS:

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

INSIGHTS

- ➢ The count of unique customers across the country is dominated by big metropolitan cities like São Paulo, Rio de Janeiro, Belo Horizonte.
- ➢ This basically indicates that big states are volume drivers whereas smaller but fast-growing states draw growth opportunities for the business.

RECOMMENDATIONS

- ➢ For top states like SP, RJ and MG, Target should push loyalty programs, subscription models, and faster delivery guarantees.
- ➢ For states with least customers, Target should analyze logistics functionality and offer introductory free shipping or regionalized discounts to build trust in first-time buyers.

11. <u>% INCREASE IN COST OF ORDERS FROM YEAR 2017-18 INCLUDING MONTHS FROM JAN-AUG ONLY:</u>

The % increase in the cost of orders within the years 2017 and 2018 including the months from January to August is displayed below with the help of a table and an SQL query.

QUERY:

```
WITH yearly_costs AS (
  SELECT
    EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
    SUM(p.payment_value) AS total_cost
  FROM `scaler-dsml-sql-465408.target.orders` AS o
  JOIN `scaler-dsml-sql-465408.target.payments` AS p
  USING(order_id)
  WHERE EXTRACT(MONTH FROM order_purchase_timestamp) BETWEEN 1 AND 8
  GROUP BY order_year
  HAVING order_year IN(2017,2018))
SELECT
ROUND(MAX(CASE WHEN order_year = 2017 THEN total_cost END),2) AS cost_2017,
ROUND(MAX(CASE WHEN order_year = 2018 THEN total_cost END),2) AS cost_2018,
ROUND(((MAX(CASE WHEN order_year = 2018 THEN total_cost END) -
        MAX(CASE WHEN order_year = 2017 THEN total_cost END)) /
        MAX(CASE WHEN order_year = 2017 THEN total_cost END)) * 100,2) AS pct_increase
FROM yearly_costs;
```

RESULTS:

Job information | Results | Visualization | JSON

Row	cost_2017	cost_2018	pct_increase
1	3669022.12	8694733.84	136.98

INSIGHTS

- ➢ Between Jan–Aug 2017 and Jan–Aug 2018, the total cost of orders increased from BRL 3.67M to BRL 8.69M, which is a +136.98% hike.
- ➢ This indicates that Target experienced more than double the sales value in just one year during the same seasonal window.
- ➢ The growth is not just marginal but explosive, implying that customers were not only returning but also spending more, possibly linked to better marketing campaigns, wider seller participation, or improved logistics.

RECOMMENDATIONS

- ➢ Target can identify which states/cities contributed the most to this surge and invest more in targeted promotions there.
- ➢ They should ensure this growth is sustainable by focusing on customer retention programs through loyalty discounts, referral schemes, or subscription benefits.
- ➢ Since revenue per order seems to be increasing, they can consider introducing more premium/high-margin product categories to capture additional value.

12. TOTAL AND AVERAGE VALUE OF ORDER PRICE FOR EACH STATE:

The total and average price of orders according to each state is presented below with the help of a table and an SQL query.

QUERY:

```
SELECT
DISTINCT c.customer_state,
ROUND(SUM(p.payment_value),1) AS total_cost,
ROUND(AVG(p.payment_value),1) AS average_cost
FROM `scaler-dsml-sql-465408.target.customers` AS c
JOIN `scaler-dsml-sql-465408.target.orders` AS o
USING(customer_id)
JOIN `scaler-dsml-sql-465408.target.payments` AS p
USING(order_id)
WHERE o.order_purchase_timestamp IS NOT NULL
GROUP BY c.customer_state
ORDER BY total_cost DESC;
```

RESULTS:

Row	customer_state	total_cost	average_cost
1	SP	5998227.0	137.5
2	RJ	2144379.7	158.5
3	MG	1872257.3	154.7
4	RS	890898.5	157.2
5	PR	811156.4	154.2
6	SC	623086.4	166.0
7	BA	616645.8	170.8
8	DF	355141.1	161.1
9	GO	350092.3	165.8
10	ES	325967.5	154.7

INSIGHTS

- ➢ São Paulo (SP) with R$ 5.99M total spend contributes the largest chunk of overall sales. However, its average order value (R$ 137.5) is the *lowest* among the top states which indicates that sales are volume-driven rather than value-driven.
- ➢ Rio de Janeiro (RJ), Minas Gerais (MG), and Rio Grande do Sul (RS) together contribute nearly R$ 4.9M.Their average order values (R$ 154–158) are higher than SP, showing customers here spend more per order.

RECOMMENDATIONS

- ➢ Target should focus on logistics optimization, fast delivery, and discount campaigns to retain its massive customer base.
- ➢ For SP, they can also consider bundling offers to push the average order value higher.
- ➢ Since RJ and MG are mid-high spenders, a mix of affordability + premium product ranges will work well. Targeted marketing could expand both order volume and value.

13. TOTAL AND AVERAGE ORDER FREIGHT VALUE FOR EACH STATE:

The total and average freight value applied on orders is represented below with the help of a table and an SQL query.

QUERY:

```
SELECT
DISTINCT c.customer_state,
ROUND(SUM(oi.freight_value),1) AS total_freight_value,
ROUND(AVG(oi.freight_value),1) AS average_freight_value
FROM `scaler-dsml-sql-465408.target.customers` AS c
JOIN `scaler-dsml-sql-465408.target.orders` AS o
USING(customer_id)
JOIN `scaler-dsml-sql-465408.target.order_items` AS oi
USING(order_id)
WHERE o.order_purchase_timestamp IS NOT NULL
GROUP BY c.customer_state
ORDER BY total_freight_value DESC;
```

RESULTS:

Row	customer_state	total_freight_value	average_freight_value
1	SP	718723.1	15.1
2	RJ	305589.3	21.0
3	MG	270853.5	20.6
4	RS	135522.7	21.7
5	PR	117851.7	20.5
6	BA	100156.7	26.4
7	SC	89660.3	21.5
8	PE	59449.7	32.9
9	GO	53115.0	22.8
10	DF	50625.5	21.0

INSIGHTS

➢ Sao Paulo dominates in freight costs with 718K total freight value. It's the top state aligning with its high order volume and customer base. However, its average freight per order (15.1) is lowest among top states, suggesting efficient logistics and better proximity to sellers/distribution hubs.

➢ States like PE show an average freight of 32.9, 2x higher than SP. This indicates either greater distance from major seller hubs or weaker logistics networks driving up costs.

RECOMMENDATIONS

➢ Target should expand regional distribution centers for states like PE, BA to reduce average freight costs.

➢ They can offer shipping discounts or free shipping thresholds in high freight states (PE, BA, SC) to incentivize orders. This can balance out cost sensitivity in these regions.

14. <u>ACTUAL DELIVERY TIME TAKEN V/S ESTIMATED TIME FOR ORDER DELIVERY TO CUSTOMERS:</u>

The number of days taken to deliver an order and the estimated time of delivery of orders is calculated and represented below with the help of a table and an SQL query.

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 `scaler-dsml-sql-465408.target.orders`
WHERE order_purchase_timestamp IS NOT NULL
AND order_delivered_customer_date IS NOT NULL
AND order_estimated_delivery_date IS NOT NULL;
```

RESULTS:

Row	order_id	time_to_deliver	diff_estimated_d...
1	65d1e226dfaeb8cdc42f665...	35	16
2	2c45c33d2f9cb8ff8b1c86cc...	30	28
3	1950d777989f6a877539f53...	30	-12
4	bfbd0f9bdef84302105ad71...	54	-36
5	98974b076b01553d49ee64...	43	6
6	c4b41c36dd589e901f6879f...	36	14
7	d2292ff2201e74c5db154d1...	29	20
8	95e01270fcbae9863423400...	30	19
9	ed8c7b1b3eb256c70ce0c7...	44	5
10	5cc475c7c03290048eb2e7...	68	-18

INSIGHTS

- ➢ Some orders take around 29–36 days, while others extend to 54–68 days. This could be due to inconsistent logistics performance across different regions or product categories.
- ➢ Positive values indicate customers received orders earlier than the estimated date. Negative values indicate orders arrived late compared to the promised date.
- ➢ Early deliveries improve satisfaction while late deliveries can frustrate customers and reduce trust.

RECOMMENDATIONS

- ➢ Target should optimize logistics in slow regions by identifying states/cities where delivery time exceeds 40 days and work with regional carriers.
- ➢ Longer delivery times may be tied to **bulky or imported products**. Target can consider faster fulfillment for high-demand categories like electronics, apparels, etc.
- ➢ They can track average delivery days, % of late deliveries, % of early deliveries per state/month to improve SLA compliance.

15. TOP 5 STATES WITH HIGHEST AND LOWEST AVERAGE FREIGHT VALUE:

The Top 5 states with Lowest and Top 5 states with Highest average freight value are listed together within one single SQL query below and presented with the help of a table.

QUERY:

```
WITH state_freight AS (
  SELECT
    c.customer_state,
    ROUND(AVG(oi.freight_value),2) AS average_freight
  FROM `scaler-dsml-sql-465408.target.customers` AS c
  JOIN `scaler-dsml-sql-465408.target.orders` AS o
  USING(customer_id)
  JOIN `scaler-dsml-sql-465408.target.order_items` AS oi
  USING(order_id)
  WHERE order_purchase_timestamp IS NOT NULL
  GROUP BY c.customer_state)
SELECT customer_state, average_freight,'Highest' AS category
FROM state_freight
QUALIFY ROW_NUMBER() OVER (ORDER BY average_freight DESC) <= 5
UNION ALL
SELECT customer_state, average_freight,'Lowest' AS category
FROM state_freight
QUALIFY ROW_NUMBER() OVER (ORDER BY average_freight ASC) <= 5;
```

RESULTS:

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

INSIGHTS

- ➢ The lowest average freight States like SP, PR, MG, RJ, DF are Brazil’s most urbanized and economically developed regions. They have dense logistics networks, short delivery distances, and high order volumes, which reduces per-order freight costs.
- ➢ The highest average freight States like RR, PB, RO, AC, PI are more remote and geographically challenging. Longer delivery distances, weaker infrastructure, and lower order volumes more likely increase freight costs.

RECOMMENDATIONS

- ➢ Partnering with local courier services in RR, PB, RO, AC, PI to reduce last-mile costs.
- ➢ It can consider freight subsidies in high-freight regions to encourage customer adoption.
- ➢ In low-cost states (SP, RJ, MG), continue leveraging free shipping as a competitive advantage.

16. <u>TOP 5 STATES WITH HIGHEST AND LOWEST AVERAGE DELIVERY TIME:</u>

The average delivery time is basically the average time taken to deliver a product from the time of order purchase. The top 5 states with highest and the top 5 states with lowest average delivery time are presented below with the help of a table and an SQL query.

QUERY:

```
WITH avg_delivery_times AS
(SELECT c.customer_state,
 ROUND(AVG(DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp,DAY)),2)
        AS average_delivery_time
 FROM `scaler-dsml-sql-465408.target.orders` AS o
 JOIN `scaler-dsml-sql-465408.target.customers` AS c
 USING(customer_id)
 GROUP BY c.customer_state),
ranked AS
(SELECT customer_state,average_delivery_time,
  RANK() OVER(ORDER BY average_delivery_time ASC) AS low_rank,
  RANK() OVER(ORDER BY average_delivery_time DESC) AS high_rank
 FROM avg_delivery_times)
SELECT
customer_state,average_delivery_time,
CASE
WHEN low_rank <= 5 THEN 'Lowest'
WHEN high_rank <= 5 THEN 'Highest'
END AS category
FROM ranked
WHERE low_rank <=5 OR high_rank <= 5
ORDER BY average_delivery_time;
```

RESULTS:

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

INSIGHTS

- ➢ SP, PR, MG, DF and SC are the fastest states. These are mostly economically developed states with dense logistics infrastructure.
- ➢ PA, AL, AM, AP and RR show much slower deliveries as they are geographically remote with limited road access and reliance on air/river transport causes delays.

RECOMMENDATIONS

- ➢ Opening secondary distribution hubs in the North and Northeast to reduce reliance on shipping from SP.
- ➢ Partner with last-mile carriers and regional couriers specializing in difficult-to-reach areas.

17. TOP 5 STATES WITH EARLIEST DELIVERY TIME COMPARING TO THE ESTIMATED DELIVERY DATE:

The top 5 states with order delivery before estimated delivery date in the earliest time possible are represented below with the help of a table and an SQL query.

QUERY:

```
SELECT
c.customer_state,
ROUND(AVG(DATE_DIFF(order_estimated_delivery_date,
      order_delivered_customer_date,DAY)),2) AS avg_days_early
FROM `scaler-dsml-sql-465408.target.orders` AS o
JOIN `scaler-dsml-sql-465408.target.customers` AS c
USING(customer_id)
WHERE order_status = 'delivered'
GROUP BY c.customer_state
ORDER BY avg_days_early DESC
LIMIT 5;
```

RESULTS:

Row	customer_state	avg_days_early
1	AC	19.76
2	RO	19.13
3	AP	18.73
4	AM	18.61
5	RR	16.41

INSIGHTS

- ➢ AC, RO, AP, AM and RR are mostly in the North & North-West regions of Brazil, which are logistically challenging because of distance, infrastructure, and transport limitations.
- ➢ The fact that customers are receiving their orders 2–3 weeks earlier than expected suggests that sellers / marketplaces are padding estimates for these states to avoid complaints and ensure SLA compliance.
- ➢ In contrast, urban states like SP, RJ, MG usually have tighter delivery commitments, so you don't see such large differences there.

RECOMMENDATIONS

- ➢ If products are consistently reaching remote states faster than estimated, Target can consider regional warehouses or 3PL tie-ups to balance estimates and actuals.
- ➢ They can highlight "Faster than expected deliveries" as a value proposition in these regions to improve brand perception.

18. MONTH-WISE COUNT OF ORDERS PLACED USING DIFFERENT PAYMENT METHODS:

The count of orders placed each month within the data using different payment methods is represented below with the help of a table and an SQL query.

QUERY:

```
SELECT
  EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
  EXTRACT(MONTH FROM o.order_purchase_timestamp) AS order_month,
  p.payment_type,
  COUNT(DISTINCT o.order_id) AS total_orders
FROM `scaler-dsml-sql-465408.target.orders` AS o
JOIN `scaler-dsml-sql-465408.target.payments` AS p
USING(order_id)
GROUP BY order_year,order_month,payment_type
ORDER BY order_year,order_month,payment_type;
```

RESULTS:

Row	order_year	order_month	payment_type	total_orders
1	2016	9	credit_card	3
2	2016	10	UPI	63
3	2016	10	credit_card	253
4	2016	10	debit_card	2
5	2016	10	voucher	11
6	2016	12	credit_card	1
7	2017	1	UPI	197
8	2017	1	credit_card	582
9	2017	1	debit_card	9
10	2017	1	voucher	33

INSIGHTS

- ➢ UPI started showing presence in late 2016 and early 2017. By Jan 2017, UPI already had 197 orders, indicating strong customer acceptance of faster, more convenient payment options.
- ➢ Very few orders are placed via debit card and credit card has been the most widely used payment method by Brazilian customers.
- ➢ Voucher-based orders do appear, but they are relatively less compared to credit card/ UPI.

RECOMMENDATIONS

- ➢ Target can partner with UPI providers and run "instant pay, instant cashback" promotions to boost adoption.
- ➢ They can build a time-series analysis of payment adoption across 2016–2018. If UPI continues to grow month-over-month, consider making it the primary promoted option alongside credit cards.
- ➢ Since debit card usage is extremely low, resources (e.g., marketing or integrations) may be better spent on UPI and credit card innovations.

19. COUNT OF ORDERS PLACED BASED ON PAYMENT INSTALLMENTS BEING PAID BY THE CUSTOMERS:

The total number of orders placed based on the number of payment installments is shown below with the help of a table and an SQL query.

QUERY:

```
SELECT
payment_installments,
COUNT(DISTINCT order_id) AS total_orders
FROM `scaler-dsml-sql-465408.target.payments`
WHERE payment_value > 0
GROUP BY payment_installments
ORDER BY payment_installments;
```

RESULTS:

Row	payment_installments	total_orders
1	0	2
2	1	49057
3	2	12389
4	3	10443
5	4	7088
6	5	5234
7	6	3916
8	7	1623
9	8	4253
10	9	644

INSIGHTS

- 49,057 orders (approx 60%) are completed with 1 installment → customers prefer paying in full rather than splitting payments. This shows strong consumer trust and purchasing power.
- Approximately 30% of orders use multiple installments. Popular installment choices are 2–4 installments (12,389 + 10,443 + 7,088 orders).
- Beyond 6 installments, the usage drops sharply showing that fewer customers prefer long-term financing.

RECOMMENDATIONS

- Since upfront payment is already the norm, Target can add small incentives (cashback, loyalty rewards, discount coupons) to strengthen this.
- 2–5 installments have the second-highest adoption, they can offer interest-free EMI plans here to attract cost-sensitive buyers.
- For rare long-term installment users, partnering with banks/fintechs to outsource risk while still keeping it available for niche customers will help.

20. WEEKDAYS RECORDING THE MOST NUMBER OF ORDERS PLACED BY THE CUSTOMERS:

The count of orders placed on each day ordered by the most number of orders is listed below with the help of a table and an SQL query. The numbers in the 'day of week' column represent the following:

1 - Sunday, 2 - Monday, 3 - Tuesday, 4 - Wednesday, 5 - Thursday, 6 - Friday, 7 - Saturday.

QUERY:

```
SELECT
  EXTRACT(DAYOFWEEK FROM order_purchase_timestamp) AS day_of_week,
  COUNT(order_id) AS total_orders
FROM `scaler-dsml-sql-465408.target.orders`
WHERE order_status = 'delivered'
GROUP BY day_of_week
ORDER BY total_orders DESC;
```

RESULTS:

Row	day_of_week	total_orders
1	2	15701
2	3	15503
3	4	15076
4	5	14323
5	6	13685
6	1	11635
7	7	10555

INSIGHTS

- Monday is the busiest day with ~15.7k orders, closely followed by Tuesday and Wednesday. This indicates customers tend to purchase more at the beginning of the work week, possibly after browsing or adding items to cart during weekends.
- Saturday is the slowest day with approx 10.5k orders, while Sunday is slightly higher (approx 11.6k). Customers may be less active online during weekends, preferring leisure or outdoor activities.

RECOMMENDATIONS

- Target can schedule flash sales or targeted promotions on weekends (Sat–Sun) to boost low activity days.
- Leveraging Monday peak traffic by launching new product drops or limited-time offers will also be proficient.
- They can send cart recovery reminders on Sunday evenings, nudging customers to complete purchases on Monday when they are most active.

- **CONCLUSION:**

The analysis of customer orders, payments, delivery times, and logistics across different states reveals valuable insights into shopping behavior and operational performance. We observed that order volumes peak at the beginning of the week, while weekends remain underutilized, presenting opportunities for targeted promotions. Payment behavior shows a heavy reliance on credit cards and single-installment purchases, though multi-installment methods still play a significant role in high-value transactions. Delivery performance varies widely across states, with some regions experiencing consistently longer lead times, while others receive orders significantly earlier than estimated, highlighting operational inefficiencies as well as areas of strength. Freight costs also differ by geography, impacting profitability and logistics strategy. Overall, the project demonstrates that customer demand patterns, payment preferences, and delivery performance are deeply influenced by both temporal and regional factors. By aligning marketing, payment offerings, and logistics strategies with these insights, Target can enhance customer satisfaction, optimize costs, and strengthen its competitive edge in the e-commerce marketplace.

- **RECOMMENDATIONS / SUGGESTIONS:**

Target should focus on improving logistics in states with longer delivery times through better regional warehousing and last-mile partnerships, while recalibrating estimated delivery windows in states where deliveries arrive much earlier to enhance customer trust. Weekend order volumes are comparatively low, so Target can introduce exclusive promotions and campaigns to drive weekend sales. Since credit cards and installments dominate payments, expanding flexible installment options and offering targeted rewards can boost high-value transactions, while partnerships with banks and fintechs can encourage adoption of underused methods. Additionally, freight costs can be optimized in high-cost states through route optimization and localized fulfillment centers. Overall, aligning delivery accuracy, payment flexibility, and marketing strategies with customer behavior will enhance satisfaction, reduce costs, and strengthen Target’s competitive position.