"Operational Insights and Performance Analysis of Target's Retail Operations in Brazil: A Data-Driven Study of Order Trends and Customer Behaviour"

Project Description

This project focuses on analyzing Target's operations in Brazil using a dataset derived from Target's database, which contains 100,000 orders placed between 2016 and 2018. The data is organized into several tables, each providing detailed information on various aspects of the business. These tables cover key areas such as order status, pricing strategies, payment and freight performance, customer demographics, product attributes, and customer satisfaction. By querying and analyzing these structured tables, the project aims to uncover valuable trends, assess operational efficiency, and identify opportunities to enhance customer experience and optimize business performance in the Brazilian market.

Problem Statement

As a data analyst/scientist at Target, my task is to analyze the provided dataset to extract valuable insights that can drive business decisions and offer actionable recommendations. This analysis will involve examining key trends, customer behaviors, sales patterns, and potential areas for optimization in various aspects of Target's operations. The goal is to leverage data to improve overall efficiency, customer satisfaction, and profitability.

Target's database consists of following tables:

1. customers

- **customer_id**: ID of the consumer who made the purchase
- **customer_unique_id**: Unique ID of the consumer
- **customer_zip_code_prefix**: Zip Code of consumer's location
- **customer_city**: Name of the city from where the order is made
- **customer_state**: State Code from where the order is made (e.g., São Paulo - SP)

2. sellers

- **seller_id**: Unique ID of the seller registered
- **seller_zip_code_prefix**: Zip Code of the seller's location
- **seller_city**: Name of the city of the seller
- **seller_state**: State Code (e.g., São Paulo - SP)

3. items

- **order_id**: Unique ID of the order made by the consumers
- **order_item_id**: Unique ID given to each item ordered in the order
- **product_id**: Unique ID given to each product available on the site
- **seller_id**: Unique ID of the seller registered in Target
- **shipping_limit_date**: Date before which the ordered product must be shipped
- **price**: Actual price of the products ordered
- **freight_value**: Price rate at which a product is delivered from one point to another

4. geolocations

- **geolocation_zip_code_prefix**: First 5 digits of Zip Code
- **geolocation_lat**: Latitude of the location
- **geolocation_lng**: Longitude of the location
- **geolocation_city**: City of the location
- **geolocation_state**: State of the location

5. payments

- **order_id**: Unique ID of the order made by the consumers
- **payment_sequential**: Sequence of payments made in case of EMI
- **payment_type**: Mode of payment used (e.g., Credit Card)
- **payment_installments**: Number of installments in case of EMI purchase
- **payment_value**: Total amount paid for the purchase order

6. orders

- **order_id**: Unique ID of the order made by the consumers
- **customer_id**: ID of the consumer who made the purchase
- **order_status**: Status of the order (e.g., delivered, shipped)
- **order_purchase_timestamp**: Timestamp of the purchase
- **order_delivered_carrier_date**: Delivery date at which carrier made the delivery
- **order_delivered_customer_date**: Date at which customer received the product

- **order_estimated_delivery_date**: Estimated delivery date of the products

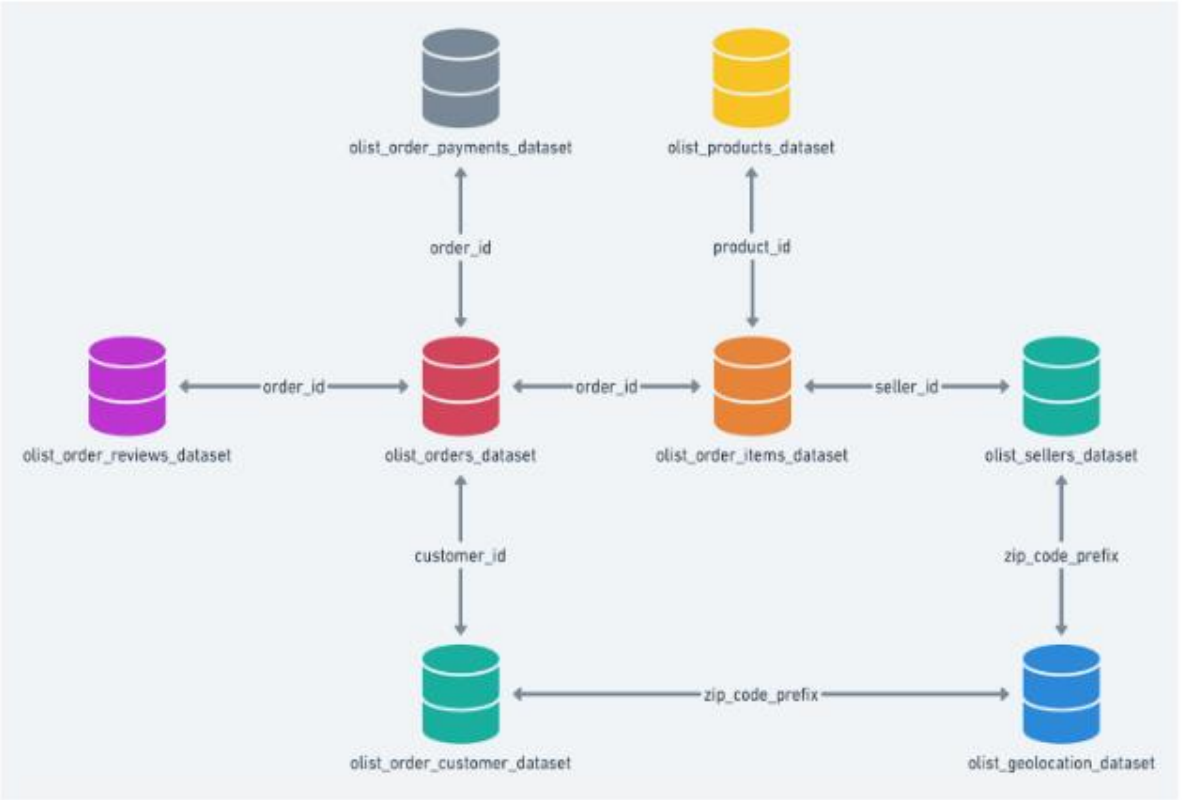
7. reviews

- **review_id**: ID of the review given on the product ordered by the order ID
- **order_id**: Unique ID of the order made by the consumers
- **review_score**: Review score given by the customer for each order (scale of 1-5)
- **review_comment_title**: Title of the review
- **review_comment_message**: Review comments posted by the consumer for each order
- **review_creation_date**: Timestamp of when the review was created
- **review_answer_timestamp**: Timestamp of when the review was answered

8. products

- **product_id**: Unique identifier for the product
- **product_category_name**: Name of the product category
- **product_name_length**: Length of the string for the name of the product
- **product_description_length**: Length of the description of the product
- **product_photos_qty**: Number of photos available for the product
- **product_weight_g**: Weight of the product in grams
- **product_length_cm**: Length of the product in cm
- **product_height_cm**: Height of the product in cm
- **product_width_cm**: Width of the product in cm

Dataset Schema



Exploratory Data Analysis (EDA)

- **Finding Data Types of all columns in a database table**

```
SELECT column_name,DATA_TYPE
from Target_Brazil.INFORMATION_SCHEMA.COLUMNS
where table_name = 'geolocation';
```

Query results

JOB INFORMATION | RESULTS | CHART | 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

- **Time range between which the orders were placed**

```
select min(order_purchase_timestamp) AS first_order_date,
max(order_purchase_timestamp) AS last_order_date
from `Target_Brazil.orders`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON

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

- The dataset includes transaction data from the first order placed on **September 4, 2016**, through to the last order on **October 17, 2018**.

Customer Behaviour Analysis

- **Query to get the highest number of customers by city or state.**

```
SELECT customer_state, customer_city, COUNT(DISTINCT customer_id) AS num_customers
FROM `Target_Brazil.customers`
GROUP BY customer_state, customer_city
ORDER BY num_customers DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	customer_state	customer_city	num_customers
1	SP	sao paulo	15540
2	RJ	rio de janeiro	6882
3	MG	belo horizonte	2773
4	DF	brasilia	2131
5	PR	curitiba	1521

- The city of São Paulo, in the state of São Paulo, has the highest number of customers making purchases at Target.

- **Query to calculate the total amount spent by each customer over time.**

```
SELECT o.customer_id, SUM(payment_value) AS total_spent
FROM `Target_Brazil.payments` p
join `Target_Brazil.orders` o
on p.order_id=o.order_id
GROUP BY customer_id
ORDER BY total_spent DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	customer_id	total_spent
1	1617b1357756262bfa56ab541...	13664.08
2	ec5b2ba62e57434238687163...	7274.88
3	c6e2731c5b391845f6800c974...	6929.31
4	f48d464a0baaea338cb25f816...	6922.21
5	3fd6777bbce08a352fddd04e4...	6726.66

➢ The customer with ‘customer_id’ ‘1617b’ has made the highest total spend on purchases at Target.

- **Query for calculating the number of orders per customer.**

```
SELECT customer_id, COUNT(DISTINCT order_id) AS num_orders
FROM `Target_Brazil.orders`
GROUP BY customer_id
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	customer_id	num_orders
1	8886130db0ea6e9e70ba0b03d...	1
2	b2191912d8ad6eac2e4dc3b6e...	1
3	622e13439d6b5a0b486c4356...	1
4	b6f6cbfc126f1ae6723fe2f9b37...	1
5	b106b360fe2ef8849fbbd056f7...	1

➢ Mostly every customer made a single purchase in Target.

Products and Orders Analysis

- **Query to identify the most frequently purchased products based on the number of orders.**

```
SELECT product_id, COUNT(order_item_id) AS num_orders
FROM `Target_Brazil.items`
GROUP BY product_id
ORDER BY num_orders DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	product_id	num_orders
1	aca2eb7d00ea1a7b8ebd4e683...	527
2	99a4788cb24856965c36a24e3...	488
3	422879e10f46682990de24d77...	484
4	389d119b48cf3043d311335e4...	392
5	368c6c730842d78016ad8238...	388

- The product with 'product_id' 'aca2eb' is the most frequently purchased item by customers.

- **Query to calculate the total revenue generated by each product based on its price.**

```
SELECT product_id, ROUND(SUM(price),2) AS total_revenue
FROM `Target_Brazil.items`
GROUP BY product_id
ORDER BY total_revenue DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	product_id	total_revenue
1	bb50f2e236e5eea0100680137...	63885.0
2	6cdd53843498f928905446678...	54730.2
3	d6160fb7873f184099d9bc95e...	48899.34
4	d1c427060a0f73f6b889a5c7c...	47214.51
5	99a4788cb24856965c36a24e3...	43025.56

➢ The product with product_id 'bb50f' generated the highest revenue for Target.

- **Query to analyze order frequency by product category**

```
SELECT product_category, COUNT(order_item_id) AS num_orders
FROM `Target_Brazil.products` p
JOIN `Target_Brazil.items` i
ON p.product_id = i.product_id
GROUP BY product_category
ORDER BY num_orders DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	product_category	num_orders
1	bed table bath	11115
2	HEALTH BEAUTY	9670
3	sport leisure	8641
4	Furniture Decoration	8334
5	computer accessories	7827

➢ The product category "bed, table, bath" received the highest number of orders from customers at Target.

Seller Performance Analysis

- **Query to calculate the number of orders for each seller.**

```
SELECT seller_id, COUNT(DISTINCT order_id) AS num_orders
FROM `Target_Brazil.items`
GROUP BY seller_id
ORDER BY num_orders DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	seller_id	num_orders
1	6560211a19b47992c3666cc44...	1854
2	4a3ca9315b744ce9f8e937436...	1806
3	cc419e0650a3c5ba77189a188...	1706
4	1f50f920176fa81dab994f9023...	1404
5	da8622b14eb17ae2831f4ac5b...	1314

➢ The seller with ‘seller_id’ ‘6560211’ made the highest number of product sales at Target.

- **Query to calculate the average delivery time by a seller**

```
SELECT i.seller_id,AVG(DATE_DIFF(o.order_delivered_customer_date,
       o.order_purchase_timestamp, DAY)) AS average_delivery_time
FROM `Target_Brazil.orders` o
JOIN `Target_Brazil.items` i
ON o.order_id = i.order_id
WHERE o.order_delivered_customer_date IS NOT NULL
      AND o.order_purchase_timestamp IS NOT NULL
GROUP BY i.seller_id
ORDER BY average_delivery_time
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JS

Row	seller_id	average_delivery_time
1	6561d6bf844e464b401944269...	1.0
2	5e063e85d44b0f5c3e6ec31311...	1.0
3	702835e4b785b67a084280efc...	1.0
4	96f7c797de9ca20efbe14545be...	1.0
5	139157dd4daa45c25b0807ffff...	1.0

➢ The top 5 sellers at Target take an average of 1 day to deliver products to customers.

- **Query for calculating the average review score (review_score) for each seller**

```
SELECT i.seller_id, AVG(r.review_score) AS avg_review_score
FROM `Target_Brazil.items` i
JOIN `Target_Brazil.reviews` r
ON i.order_id = r.order_id
GROUP BY i.seller_id
ORDER BY avg_review_score DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	seller_id	avg_review_score
1	77128dec4bec4878c37ab7d61...	5.0
2	cf6f6bc4df3999b9c6440f124f...	5.0
3	6c9875b2f94ba781186f0c1ae...	5.0
4	702835e4b785b67a084280efc...	5.0
5	48efc9d94a9834137efd9ea76...	5.0

➢ The majority of sellers at Target receive a 5-star rating.

Geographic and Location-Based Analysis

- **Aggregate sales by customer region, and order data.**

```
SELECT customer_state, ROUND(SUM(payment_value),2) AS total_sales
FROM `Target_Brazil.orders` o
JOIN `Target_Brazil.payments` p
ON o.order_id = p.order_id
JOIN `Target_Brazil.customers` c
ON o.customer_id = c.customer_id
GROUP BY customer_state
ORDER BY total_sales DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	customer_state	total_sales
1	SP	5998226.96
2	RJ	2144379.69
3	MG	1872257.26
4	RS	890898.54
5	PR	811156.38

➢ The state of São Paulo generates the most revenue for Target.

- **Count of orders by customer region, and order data.**

```
SELECT customer_state, COUNT(DISTINCT order_id) AS num_orders
FROM `Target_Brazil.customers` c
JOIN `Target_Brazil.orders` o
ON c.customer_id=o.customer_id
GROUP BY customer_state
ORDER BY num_orders DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

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

➢ The state of São Paulo has the highest number of orders at Target.

Payment and Financial Analysis

- **Query to calculate the distribution of payment types used by customers (e.g., Credit Card, Debit Card, etc.).**

```
SELECT payment_type, COUNT(DISTINCT order_id) AS num_orders
FROM `Target_Brazil.payments`
GROUP BY payment_type
ORDER BY num_orders DESC;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	payment_type	num_orders
1	credit_card	76505
2	UPI	19784
3	voucher	3866
4	debit_card	1528
5	not_defined	3

➢ Credit card is the most commonly used payment method by customers at Target.

- **Query to analyze the average number of installments used by customers.**

```
SELECT ROUND(AVG(payment_installments)) AS avg_installments
FROM `Target_Brazil.payments`
WHERE payment_installments IS NOT NULL;
```

Query results

JOB INFORMATION

Row	avg_installments
1	3.0

➢ On average, most customers choose to pay for their purchase in 3 installments.

- **Query to calculate the total revenue from each payment type.**

```
SELECT payment_type, ROUND(SUM(payment_value),2) AS total_revenue
FROM `Target_Brazil.payments`
GROUP BY payment_type
ORDER BY total_revenue DESC;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	payment_type	total_revenue
1	credit_card	12542084.19
2	UPI	2869361.27
3	voucher	379436.87
4	debit_card	217989.79
5	not_defined	0.0

➢ The majority of revenue generated by customers comes from purchases made with credit cards.

Product Review and Sentiment Analysis

- **Query to calculate the average review score for each product based on customer reviews**

```
SELECT i.product_id, AVG(r.review_score) AS avg_review_score
FROM `Target_Brazil.items` i
JOIN `Target_Brazil.reviews` r
ON i.order_id = r.order_id
GROUP BY i.product_id
ORDER BY avg_review_score DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART

Row	product_id	avg_review_score
1	680cc8535be7cc69544238c1d...	5.0
2	46fce52cef5caa7cc225a5531c...	5.0
3	2e8316b31db34314f393806fd...	5.0
4	1716ea399ed8ee62ba811e6f5...	5.0
5	310dc32058903b6416c71faff1...	5.0

➢ The majority of products sold at Target receive a 5-star rating.

- **Query to search for basic keyword sentiment analysis by searching for positive or negative words in review_comments**

```
SELECT order_id, review_score,
       CASE
         WHEN LOWER(review_comment_title) LIKE '%good%' THEN 'positive'
         WHEN LOWER(review_comment_title) LIKE '%bad%' THEN 'negative'
         ELSE 'neutral'
       END AS sentiment
FROM `Target_Brazil.reviews`
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON

Row	order_id	review_score	sentiment
1	54d744a4410b1edccc36c6d1f...	1	neutral
2	a9a93c428c6103f2151bb63a1...	1	neutral
3	1cb796218c383fc54a6a45414...	1	neutral
4	b3feb3846bb0a8d68cd328138...	1	neutral
5	745e2506fb647deca4669e1c8...	1	neutral

➢ The majority of orders at Target receive neutral review comments.

Time Series Analysis

- **Query to identify trends in order volume or total sales over a specific time period**

```
SELECT EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
       EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
       COUNT(DISTINCT order_id) AS num_orders
FROM `Target_Brazil.orders`
GROUP BY year, month
ORDER BY year, month;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON

Row	year	month	num_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
11	2017	8	4331
12	2017	9	4285

➢ The last quarter of the year has the fewest number of orders.

- **Query to track review scores over time to identify patterns in customer satisfaction**

```
SELECT EXTRACT(MONTH FROM review_creation_date) AS month,
       ROUND(AVG(review_score),2) AS avg_review_score
FROM `Target_Brazil.reviews`
GROUP BY month
ORDER BY month;
```

Query results

JOB INFORMATION | RESULTS | CH

Row	month	avg_review_score
1	1	4.07
2	2	4.06
3	3	3.8
4	4	3.95
5	5	4.16
6	6	4.17
7	7	4.25
8	8	4.21
9	9	4.18
10	10	4.18
11	11	4.1
12	12	3.92

➢ In the month of March, products receive the lowest review scores.

Cross – Selling and Market Basket Analysis

- **Query to analyze which products are commonly purchased together**

```
WITH ProductPairs AS (
  SELECT i1.order_id, i1.product_id AS product1, i2.product_id AS product2
  FROM `Target_Brazil.items` i1
  JOIN `Target_Brazil.items` i2
  ON i1.order_id = i2.order_id AND i1.product_id != i2.product_id
)
SELECT product1, product2, COUNT(*) AS co_occurrences
FROM ProductPairs
GROUP BY product1, product2
ORDER BY co_occurrences DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	product1	product2	co_occurrences
1	05b515fdc76e888aada3c6d66c...	270516a3f41dc035aa87d2202...	100
2	270516a3f41dc035aa87d2202...	05b515fdc76e888aada3c6d66c...	100
3	36f60d45225e60c7da4558b07...	e53e557d5a159f5aa2c5e995df...	48
4	e53e557d5a159f5aa2c5e995df...	36f60d45225e60c7da4558b07...	48
5	a9d9db064d4afd4458eb3e139f...	710b7c26b7a742f497bba45fab...	36

Seller and Delivery Analysis

- **Query to calculate the percentage of orders delivered on time**

```
SELECT s.seller_id,
       ROUND(COUNT(CASE WHEN DATE_DIFF(order_delivered_customer_date,
order_estimated_delivery_date, DAY) <= 0 THEN 1 END) /
       COUNT(*),2) AS on_time_delivery_rate
FROM `Target_Brazil.sellers` s
JOIN `Target_Brazil.items` i
ON s.seller_id=i.seller_id
JOIN `Target_Brazil.orders` o
ON o.order_id=i.order_id
GROUP BY s.seller_id
ORDER BY on_time_delivery_rate DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JS

Row	seller_id	on_time_delivery_rate
1	325f3178fb58e2a9778334621...	1.0
2	6338ea67c41078a46ad99cc00...	1.0
3	a218df5c097b0ca409cfaac14...	1.0
4	9cea4682e16655f7da87f4069...	1.0
5	48efc9d94a9834137efd9ea76...	1.0

- The majority of orders were delivered to customers on time.

Recommendations

1. **Target Marketing in High-Customer Areas**: Focus marketing efforts on the cities and states with the highest number of customers. Special promotions or personalized campaigns could help retain and increase customer loyalty in these regions.

2. **Customer Retention Strategy for High Spenders**: Implement loyalty programs or special offers for high-value customers, as they contribute heavily to overall revenue.

3. **Optimize Inventory Based on Popular Products**: Stock up on the most frequently purchased products to meet demand and avoid stockouts, especially for bestsellers.

4. **Product Pricing Strategy**: Products with high revenue generation should be analyzed for potential pricing optimization. They may offer opportunities for premium pricing, bundling, or cross-selling.

5. **Seller Training for Improved Performance**: Sellers who perform poorly could benefit from training in customer service, inventory management, or marketing to boost their order volumes and overall success.

6. **Analyze Negative Reviews**: Focus efforts on understanding and improving the products with low average review scores, particularly those that customers identify as problematic.

7. **Enhance Payment Options**: Considering that credit cards dominate as the most common payment method, consider introducing more options like digital wallets or installments to cater to different customer preferences

.

8. **Optimize Delivery Performance**: Sellers should be incentivized or penalized based on their delivery performance. Improving on-time delivery rates will enhance customer satisfaction and reduce complaints.

9. **Geographic Sales Expansion**: Consider expanding business operations or marketing in regions with lower sales, as there may be untapped potential for growth

.

10. **Leverage Cross-Selling Opportunities**: The products commonly bought together should be promoted as bundles or recommended together on the website to increase sales through cross-selling.