# Business Case: Target SQL

Scaler DS ML

Shayantan Dey

Sat 20th July

**Context:**

Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can 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.

**Dataset:**

The data is available in 8 csv files at Google Drive

1. customers.csv
2. sellers.csv
3. order_items.csv
4. geolocation.csv
5. payments.csv
6. reviews.csv
7. orders.csv
8. products.csv

The column description for these csv files is given below.

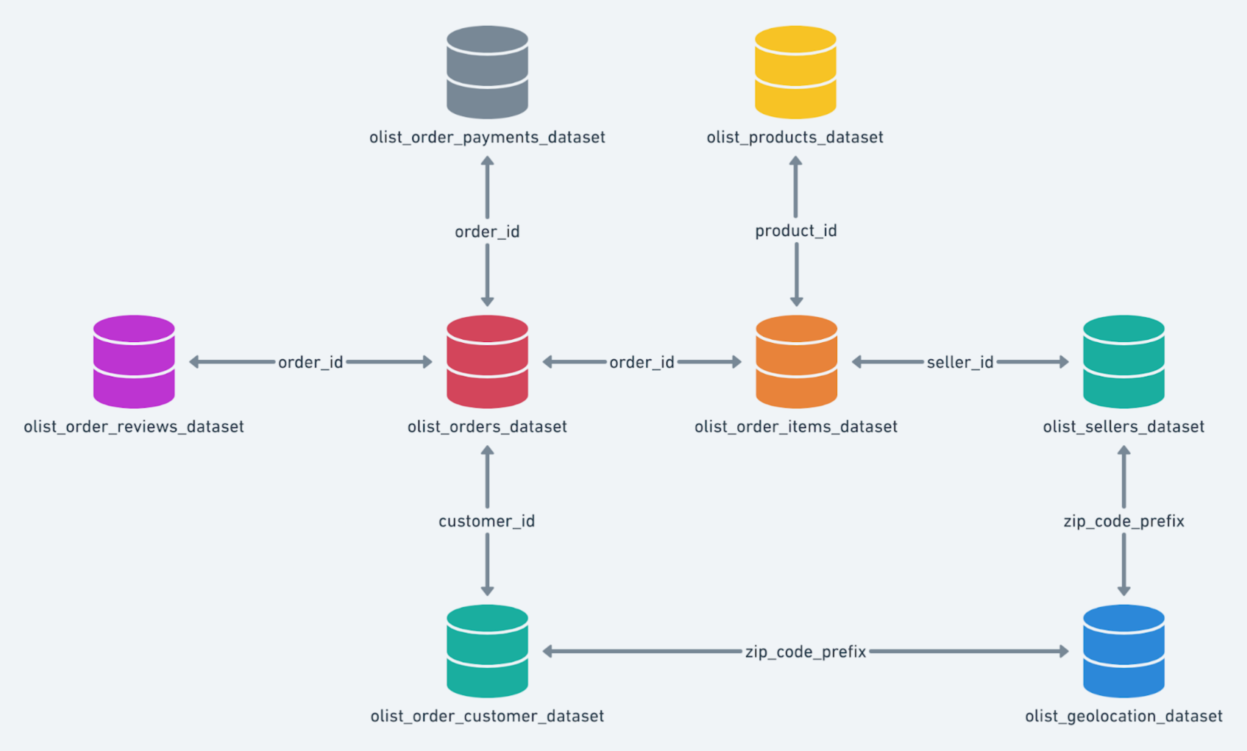
The **customers.csv** contain following features:

| Features | Description |
|---|---|
| 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 order is made |
| customer_state | State Code from where order is made (Eg. são paulo - SP) |

The **sellers.csv** contains following features:

| Features | Description |
|---|---|
| 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 (Eg. são paulo - SP) |

**Dataset schema:**



**Problem Statement:**

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

**What does 'good' look like?**

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

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

```
PRAGMA table_info(customers);
```

Table 3: 5 records

| cid | name | type | notnull | dflt_value | pk |
|---|---|---|---|---|---|
| 0 | customer_id | TEXT | 0 | NA | 0 |
| 1 | customer_unique_id | TEXT | 0 | NA | 0 |
| 2 | customer_zip_code_prefix | INTEGER | 0 | NA | 0 |
| 3 | customer_city | TEXT | 0 | NA | 0 |
| 4 | customer_state | TEXT | 0 | NA | 0 |

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

```
SELECT
    MIN(order_purchase_timestamp) AS order_start_date,
    MAX(order_purchase_timestamp) AS order_end_date,
    ROUND((julianday(MAX(order_purchase_timestamp)) -
      julianday(MIN(order_purchase_timestamp))), 2) AS order_time_range_days
FROM
    orders;
```

Table 4: 1 records

| order_start_date | order_end_date | order_time_range_days |
|---|---|---|
| 2016-09-04 21:15:19 | 2018-10-17 17:30:18 | 772.84 |

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

```
SELECT * FROM orders LIMIT 10;
```

Table 5: Displaying records 1 - 10

| order_id | customer_id | order_status | order_purchase_timestamp | order_approved_at | order_delivered_carrier_date | order_delivered_customer_date | order_estimated_delivery_d |
|---|---|---|---|---|---|---|---|
| e481f51cbdc54678b7cc49136f2d6af7 | 9ef432eb6251297304e76186b10a928d | delivered | 2017-10-02 10:56:33 | 2017-10-02 11:07:15 | 2017-10-04 19:55:00 | 2017-10-10 21:25:13 | 2017-10-18 00:00:00 |
| 53cdb2fc8bc7dce0b6741e2150273451 | b0830fb4747a6c6d20dea0b8c802d7ef | delivered | 2018-07-24 20:41:37 | 2018-07-26 03:24:27 | 2018-07-26 14:31:00 | 2018-08-07 15:27:45 | 2018-08-13 00:00:00 |
| 47770eb9100c2d0c44946d9cf07ec65d | 41ce2a54c0b03bf3443c3d931a367089 | delivered | 2018-08-08 08:38:49 | 2018-08-08 08:55:23 | 2018-08-08 13:50:00 | 2018-08-17 18:06:29 | 2018-09-04 00:00:00 |
| 949d5b44dbf5de918fe9c16f97b45f8a | f88197465ea7920adcdbec7375364d82 | delivered | 2017-11-18 19:28:06 | 2017-11-18 19:45:59 | 2017-11-22 13:39:59 | 2017-12-02 00:28:42 | 2017-12-15 00:00:00 |
| ad21c59c0840e6cb83a9ceb5573f8159 | 8ab97904e6daea8866dbdbc4fb7aad2c | delivered | 2018-02-13 21:18:39 | 2018-02-13 22:20:29 | 2018-02-14 19:46:34 | 2018-02-16 18:17:02 | 2018-02-26 00:00:00 |

| order_id | customer_id | order_status | order_purchase_timestamp | order_approved_at | order_delivered_carrier_date | order_delivered_customer_date | order_estimated_delivery_d |
|---|---|---|---|---|---|---|---|
| a4591c265e18cb1[illegible] | [illegible] | delivered | 2017-07-09 21:57:05 | 2017-07-09 22:10:13 | 2017-07-11 14:58:04 | 2017-07-26 10:57:55 | 2017-08-01 00:00:00 |
| 136cce7faa42fdb[illegible] | [illegible] | [illegible] | 2017-04-11 12:22:08 | 2017-04-13 13:25:17 | | | 2017-05-09 00:00:00 |
| 6514b8ad8028c9f[illegible] | [illegible] | [illegible] | 2017-05-16 13:10:30 | 2017-05-16 13:22:11 | 2017-05-22 10:07:46 | 2017-05-26 12:55:51 | 2017-06-07 00:00:00 |
| 76c6e866289321a[illegible] | [illegible] | delivered | 2017-01-23 18:29:09 | 2017-01-25 02:50:47 | 2017-01-26 14:16:31 | 2017-02-02 14:08:10 | 2017-03-06 00:00:00 |
| e69bfb5eb88e0ed[illegible] | [illegible] | delivered | 2017-07-29 11:55:02 | 2017-07-29 12:05:32 | 2017-08-10 19:45:24 | 2017-08-16 17:14:30 | 2017-08-23 00:00:00 |