# Target SǪL Business Case

## Why this case study?

### From company’s perspective:

* 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 different csv ﬁles:

1. customers.csv
2. geolocation.csv
3. order\_items.csv
4. payments.csv
5. reviews.csv
6. orders.csv
7. products.csv
8. sellers.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 geolocations.csv contain following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| geolocation\_zip\_code\_preﬁx | First 5 digits of Zip Code |
| geolocation\_lat | Latitude |
| geolocation\_lng | Longitude |
| geolocation\_city | City |
| geolocation\_state | State |

#### 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) |

**The order\_items.csv contain following features:**

|  |  |
| --- | --- |
| **Features** | **Description** |
| order\_id | A Unique ID of order made by the consumers |
| order\_item\_id | A Unique ID given to each item ordered in the order |
| product\_id | A Unique ID given to each product available on the site |
| seller\_id | Unique ID of the seller registered in Target |
| shipping\_limit\_date | The date before which shipping of the ordered product must be completed |
| price | Actual price of the products ordered |
| freight\_value | Price rate at which a product is delivered from one point to another |

#### The payments.csv contain following features:

|  |  |
| --- | --- |
| **Features** | **Description** |
| order\_id | A Unique ID of order made by the consumers. |
| payment\_sequential | Sequences of the payments made in case of EMI |
| payment\_type | Mode of payment used (Eg. Credit Card) |
| payment\_installments | Number of installments in case of EMI purchase. |

|  |  |
| --- | --- |
| payment\_value | Total amount paid for the purchase order |

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

|  |  |
| --- | --- |
| **Features** | **Description** |
| order\_id | A Unique ID of order made by the consumers |
| customer\_id | ID of the consumer who made the purchase |
| order\_status | Status of the order made i.e. delivered, shipped, etc. |
| 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 got the product |
| order\_estimated\_delivery\_date | Estimated delivery date of the products |

#### The reviews.csv contain following features:

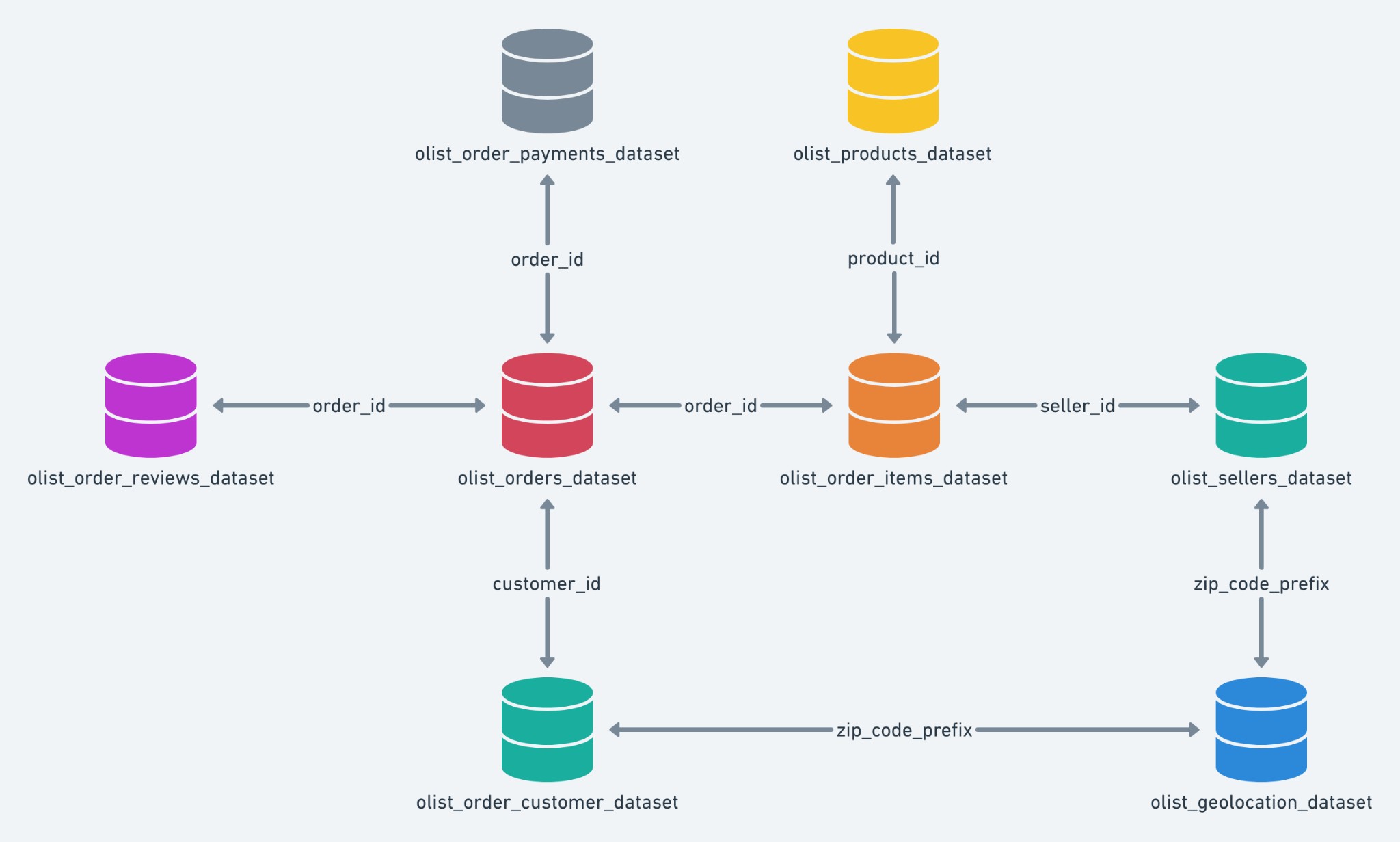
|  |  |
| --- | --- |
| **Features** | **Description** |
| review\_id | ID of the review given on the product ordered by the order id |
| order\_id | A Unique ID of order made by the consumers |
| review\_score | Review score given by the customer for each order on a scale of 1-5 |
| review\_comment\_title | Title of the review |
| review\_comment\_message | Review comments posted by the consumer for each order |
| review\_creation\_date | Timestamp of the review when it is created |
| review\_answer\_timestamp | Timestamp of the review answered |

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

|  |  |
| --- | --- |
| **Features** | **Description** |

|  |  |
| --- | --- |
| product\_id | A Unique identiﬁer for the proposed project. |
| product\_category\_name | Name of the product category |
| product\_name\_lenght | Length of the string which speciﬁes the name given to the products ordered |
| product\_description\_lenght | Length of the description written for each product ordered on the site. |
| product\_photos\_qty | Number of photos of each product ordered available on the shopping portal |
| product\_weight\_g | Weight of the products ordered in grams |
| product\_length\_cm | Length of the products ordered in centimeters |
| product\_height\_cm | Height of the products ordered in centimeters |
| product\_width\_cm | Width of the product ordered in centimeters |

**Dataset schema:**



## What does ‘good’ look like?

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

* 1. Data type of all columns in the “customers” table.

**Hint:** We want you to display the data type of each column present in the “customers” table.

* 1. Get the time range between which the orders were placed.

**Hint:** We want you to get the date & time when the ﬁrst and last orders in our dataset were placed.

* 1. Count the number of Cities and States in our dataset.

**Hint:** We want you to count the number of unique cities and states present in our dataset.

#### In-depth Exploration:

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

**Hint:** We want you to ﬁnd out if no. of orders placed has increased gradually in each month, over the past years.

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

**Hint:** We want you to ﬁnd out if the no. of orders placed are at peak during certain months.

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

**Hint:** We want you to categorize the hours of a day into the given time brackets/ intervals and ﬁnd out during which intervals the Brazilian customers usually order the most.

#### Evolution of E-commerce orders in the Brazil region:

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

**Hint:** We want you to get the no. of orders placed in each state, in each month by our customers.

* 1. How are the customers distributed across all the states?

**Hint:** We want you to get the no. of unique customers present in each state.

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

* 1. Get the % increase in the cost of orders from year 2017 to 2018 *(include months between Jan to Aug only).*

**Hint:** You can use the payment\_value column in the payments table to get the cost of orders.

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

**Hint:** We want you to fetch the total price and the average price of orders for each state.

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

**Hint:** We want you to fetch the total freight value and the average freight value of orders for each state.

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

* 1. Find the no. of days taken to deliver each order from the order’s purchase date as delivery time.

Also, calculate the difference (in days) between the estimated & actual delivery date of an order.

Do this in a single query.

**Hint:** You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

* + - **time\_to\_deliver** = order\_delivered\_customer\_date - order\_purchase\_timestamp
    - **diff\_estimated\_delivery** = order\_estimated\_delivery\_date - order\_delivered\_customer\_date
  1. Find out the top 5 states with the highest & lowest average freight value.

**Hint:** We want you to ﬁnd the top 5 & the bottom 5 states arranged in increasing order of the average freight value.

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

**Hint:** We want you to ﬁnd the top 5 & the bottom 5 states arranged in increasing order of the average delivery time.

* 1. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

You can use the difference between the averages of actual & estimated delivery date to ﬁgure out how fast the delivery was for each state.

**Hint:** Include only the orders that are already delivered.

#### Analysis based on the payments:

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

**Hint:** We want you to count the no. of orders placed using different payment methods in each month over the past years.

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

**Hint:** We want you to count the no. of orders placed based on the no. of payment installments where at least one installment has been successfully paid.