**Context**

Target is one of the world’s most recognized brands and one of America’s leading retailers. 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 business case has information of 100k orders from 2016 to 2018 made at Target in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers.

**Dataset**: <https://drive.google.com/drive/folders/1TGEc66YKbD443nslRi1bWgVd238gJCnb>

Data is available in 8 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

Each feature or columns of different CSV files are described 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 the location of the consumer. |
| customer\_city | Name of the City from where order is made. |
| customer\_state | State Code from where order is made(Ex- sao 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 location of the seller. |
| seller\_city | Name of the City of the seller. |
| seller\_state | State Code (Ex- sao 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 **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 name |
| geolocation\_state | state |

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.(Ex-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. |

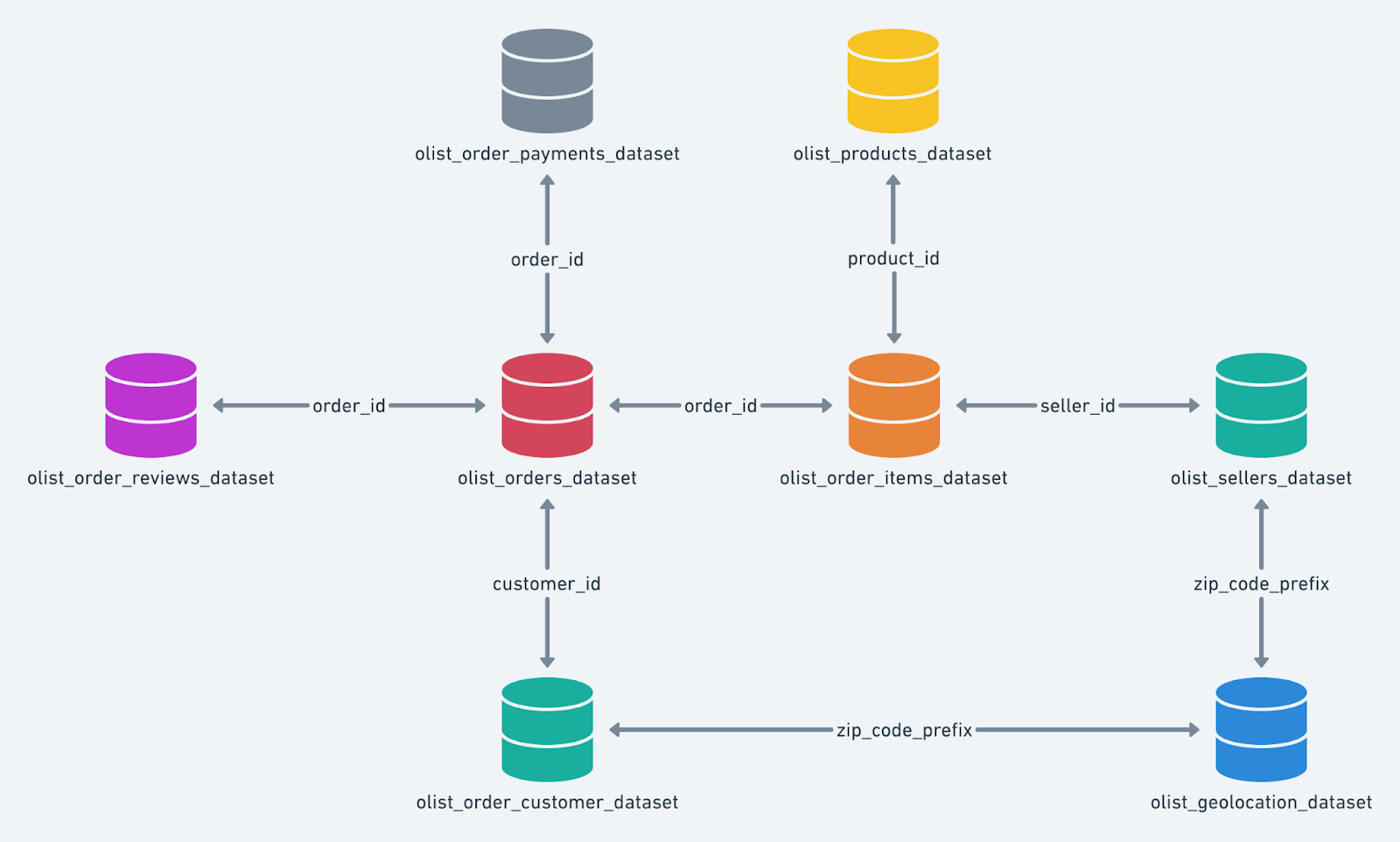
The **reviews.csv** contain following features:

|  |  |
| --- | --- |
| **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 the 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. |

**High level overview of relationship between datasets:**



Assume you are a data scientist at Target, and are given this data to analyze and provide some insights and recommendations from it.

**What does ‘good’ look like?**

1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset
   1. Data type of columns in a table
   2. Time period for which the data is given
   3. Cities and States of customers ordered during the given period
2. In-depth Exploration:
   1. Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?
   2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?
3. Evolution of E-commerce orders in the Brazil region:
   1. Get month on month orders by states
   2. Distribution of customers across the states in Brazil
4. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.
   1. Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only) - You can use “payment\_value” column in payments table
   2. Mean & Sum of price and freight value by customer state

5. Analysis on sales, freight and delivery time

1. Calculate days between purchasing, delivering and estimated delivery
2. Find time\_to\_delivery & diff\_estimated\_delivery. Formula for the same given below:
   * time\_to\_delivery = order\_purchase\_timestamp-order\_delivered\_customer\_date
   * diff\_estimated\_delivery = order\_estimated\_delivery\_date-order\_delivered\_customer\_date
3. Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery
4. Sort the data to get the following:
5. Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5
6. Top 5 states with highest/lowest average time to delivery
7. Top 5 states where delivery is really fast/ not so fast compared to estimated date

6. Payment type analysis:

1. Month over Month count of orders for different payment types
2. Count of orders based on the no. of payment installments

**Evaluation Criteria (80 points)**

1. Initial exploration of dataset like checking the characteristics of data (10 points)
2. In-depth Exploration (10 points)
3. Evolution of E-commerce orders in the Brazil region (10 points)
4. Impact on Economy (10 points)
5. Analysis on sales, freight and delivery time (10 points)
6. Payment type analysis (10 points)
7. Actionable Insights (10 points)
8. Recommendations (10 points)