

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.

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

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

Dataset: <https://drive.google.com/drive/folders/1TGEc66YKbD443nsIRi1bWgVd238gJCnb>

What 'good' looks 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 covered in the dataset
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 region, states

2. How are customers distributed 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)
 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. Create columns:
 - `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:
 1. Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5
 2. Top 5 states with highest/lowest average time to delivery
 3. 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. Distribution of payment installments and count of orders