

# CAPSTONE PROJECT 2 - IDEAS

## IDEA 1: Customer Satisfaction Prediction

### PROBLEM STATEMENT:

Many companies today measure customer satisfaction. Some of the quick ways of collecting the feedback from customers is through their reviews or by surveying them. These are very important customer service metrics, but are not typically being used to improve operations or help reduce customer churn.

Instead of waiting until the customer interaction is over for feedback/ review, companies can predict how likely an operation is to receive a good or bad rating while they are still in contact with the customer.

My main hypothesis is that the product and how the order was fulfilled influences customer review rating.

### DATASET:

Brazilian E-Commerce Public Dataset by Olist

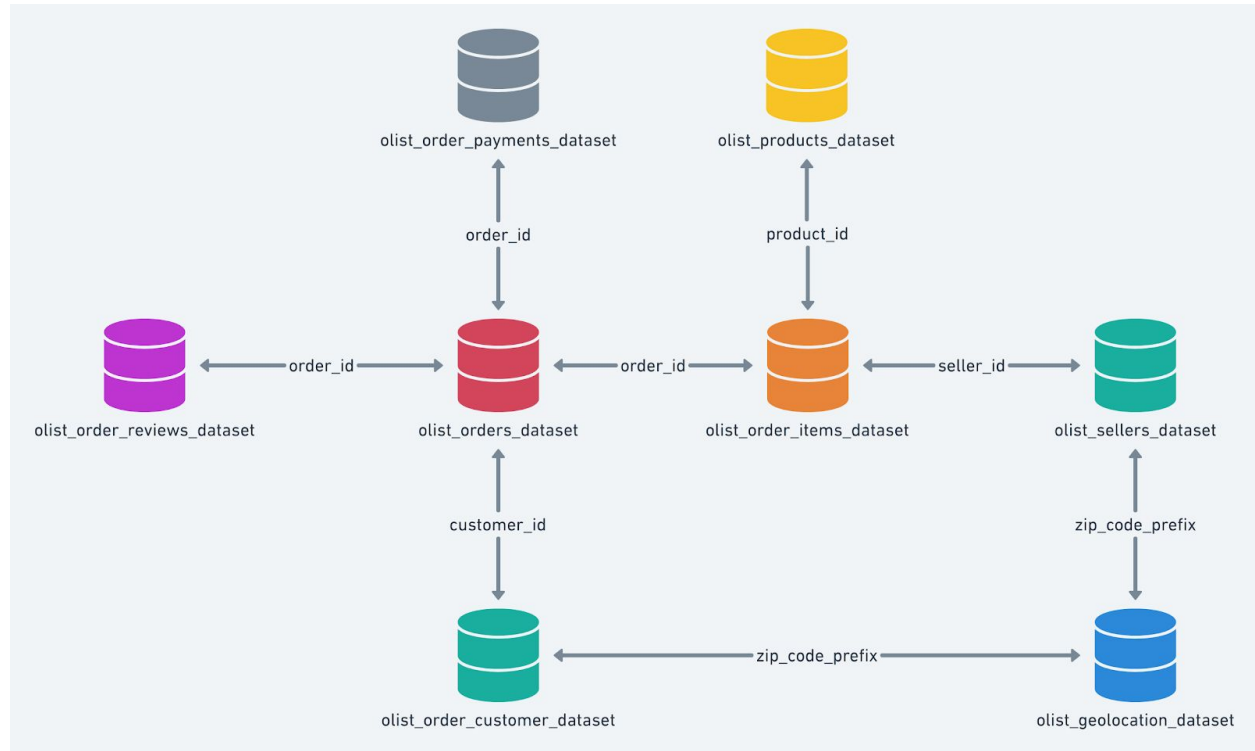
<https://www.kaggle.com/olistbr/brazilian-ecommerce>

This is a Brazilian ecommerce public dataset of orders made at Olist Store. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. It's 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.

This is real commercial data, it has been anonymised, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

## Data Schema

The data is divided in multiple datasets for better understanding and organization.



## Datasets and data descriptions:

### 1. Customers Dataset

**Customer\_id** : key to the orders dataset. Each order has a unique **customer\_id**.

**Customer\_unique\_id**: unique identifier of a customer.

**Customer\_zip\_code\_prefix**: first five digits of customer zip code

**Customer\_city**: customer city name

**Customer\_state**: customer state

### 2. Geolocation Dataset

**Geolocation\_zip\_code\_prefix**: first 5 digits of zip code

**Geolocation\_lat**: latitude

**Geolocation\_lng**: longitude

**Geolocation\_city**: city name

**Geolocation\_state**: state

### 3. Order Items Dataset

Order\_id: order unique identifier

Order\_item\_id: sequential number identifying number of items included in the same order.

Product\_id: product unique identifier

Seller\_id: seller unique identifier

Shipping\_limit\_date: Shows the seller shipping limit date for handling the order over to the logistic partner.

Price: item price

Freight\_value: item freight value item (if an order has more than one item the freight value is splitted between items)

### 4. Payments Dataset

Payment\_sequential: a customer may pay an order with more than one payment method. If he does so, a sequence will be created to accommodate all payments.

Payment\_type : method of payment chosen by the customer.

Payment\_installments : number of installments chosen by the customer.

Payment\_value: transaction value.

### 5. Order Reviews Dataset

review\_comment\_title: Comment title from the review left by the customer, in Portuguese.

Review\_comment\_message: Comment message from the review left by the customer, in Portuguese.

Review\_creation\_date : Shows the date in which the satisfaction survey was sent to the customer.

Review\_answer\_timestamp: Shows satisfaction survey answer timestamp.

### 6. Order Dataset :

Order\_approved\_at : Shows the payment approval timestamp.

Order\_delivered\_carrier\_date: Shows the order posting timestamp. When it was handled to the logistic partner.

Order\_delivered\_customer\_date : Shows the actual order delivery date to the customer.

Order\_estimated\_delivery\_date: Shows the estimated delivery date that was informed to customer at the purchase moment.

## 7. Products Dataset:

product\_photos\_qty: number of product published photos  
Product\_weight\_g: product weight measured in grams.  
Product\_length\_cm: product length measured in centimeters.  
Product\_height\_cm: product height measured in centimeters.  
Product\_width\_cm: product width measured in centimeters.

## 8. Sellers Dataset

Seller\_id: seller unique identifier  
Seller\_zip\_code\_prefix: first 5 digits of seller zip code  
Seller\_city: seller city name  
Seller\_state :seller state

## 9. Category Name Translation

Product\_category\_name: category name in Portuguese  
Product\_category\_name\_english : category name in English

## PROPOSED METHODOLOGY:

Exploratory data analysis will answer some of the questions like:

1. Discovering the products categories that are more prone to customer dissatisfaction
2. Which City has more customers
3. How long it took orders to be delivered

## DELIVERABLE:

The project would be delivered in the form of a presentation with all visualizations and results. In addition, python jupyter notebooks and a report that would detail the approach used including the steps from Data wrangling to the machine learning algorithm used.

## **Idea 2: Customer segmentation**

Data set: E-Commerce Data-Actual transactions from UK retailer –

<https://www.kaggle.com/carrie1/ecommerce-data>

## **Idea 3: Attribute extraction from e-commerce product description -NLP**

This project is to see how one can use product title and description on e-commerce websites to extract different attributes of the product.

This is a very fundamental problem in e-commerce which has widespread implications for Product Search (search filters), Product Matching (matching same items from different sellers), Product Grouping (grouping items by variants such as size and color), Product Graph (relationship between products based on attributes) and so on.

Datasets:

<http://snap.stanford.edu/data/amazon-meta.html>

<https://data.world/promptcloud/fashion-products-on-amazon-com>

<https://developer.ebay.com/>

<https://data.world/datafiniti/amazon-and-best-buy-electronics>

## **Idea 4 : anticipatory shipping**

Generate transaction forecasts for individual customers.

Use stats such as previous orders, basket or shopping cart contents, customer wish-lists and even popular regional listings

Data set: E-Commerce Data-Actual transactions from UK retailer –

<https://www.kaggle.com/carrie1/ecommerce-data>