Olist Retention Analysis Project Description

Problem: What drives customer retention at Olist and who are its most valuable customers?

Retention analytics is at the core of all digital B2C businesses. In many cases, it is crucial to ensure the economic feasibility of a product/service. This is especially the case for e-commerce businesses, who want the customers that they have acquired to make as many orders as possible to increase their lifetime value and drive up the ROI on their acquisition, this is the volume effect. Long-term customer retention allows for another value leverage, the price effect, as loyal customers tend to be less price-sensitive (Kanghyun & Thanh, 2011).

Because of these factors, it is interesting and important for digital marketplaces to analyze their customer retention data and identify opportunities for minimizing churn and improving their unit economics. It also allows the firm to perform a more granular customer segmentation and identify its most valuable clients. We aim to perform this analysis and answer these questions for the Brazilian e-commerce aggregator Olist.

Dataset: Brazilian E-Commerce Public Dataset by Olist

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

Olist is an e-commerce aggregator which serves as a platform for connecting merchants and their products to customers in Brazil. They have published a dataset of information on all 100k orders placed through their service between 2016 and 2018. This dataset is divided into multiple sub-datasets, including customers, geolocation, order items, order payments, order reviews, orders, products, and sellers. These are all in the comma-separated value .csv format.

General Plan of Attack:

First, we will perform initial exploratory analysis of the dataset and classify Olist customers as either retained or lost between 2016 and 2018. Then we will use causal inference techniques to identify the main factors that cause churn and retention. After understanding what drives retention, we will predict Customer Lifetime Value (CLTV) for Olist customers using logistic regression models. By the end of this analysis, we will have identified not only which factors make customers likely to stay Olist customers or leave, but also which customers Olist should specifically aim to retain to maximize the ROI on the acquisition cost. This will position us to recommend interventions such as targeted advertisement for customers with potential high CLTV.

Group Members: Amaury Laurent apl2163; Daiyang Li dl3393; Charlotte Liao el3073; Jonathan Jasper jfj2121; Nicholas Etz npe2105