

Capstone Project Proposal: RFM Analysis of Olist Brazilian E-Commerce Transactions

In this project, I propose to use this dataset:

<https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce> to identify distinct customer segments using RFM analysis. The ultimate goal of this analysis is to provide actionable insights to improve retention and marketing ROI. In the process we will endeavor to validate the generated clusters using business logic and statistical metrics.

The data set describes over 100k transactions occurring between 2016 and 2018. Its features allow 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. There is also a geolocation dataset that relates Brazilian zip codes to lat/long coordinates. This is both useful and also a potential challenge as the data must be cleaned appropriately for our analysis. Obviously, the potential for useful features to fuel actionable insights makes the effort worthwhile.