Consumer Data Research Centre

An ESRC Data Investment

SHOP DIRECT

Analysing customers who use discounts

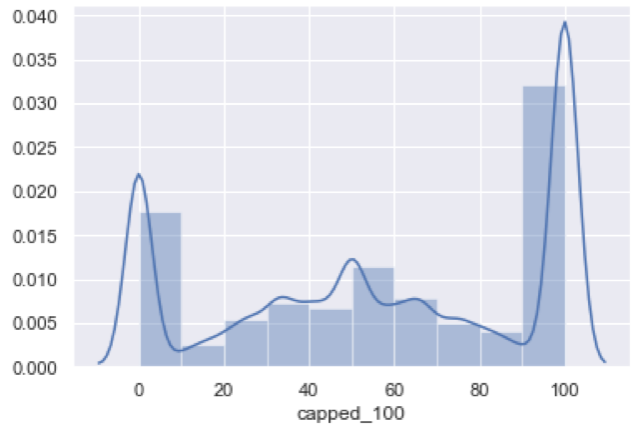
Marc Whalley[1], Cagatay Turkay[1] and Nicola Dunford[2]
[1]City University of London, [2]Shop Direct

Project Background

This CDRC placement was placed in a company called Shop Direct (SD). SD is a large online retailer that used to have a high street presence (Littlewoods) and has diversified its offer to customers by higher end product ranges such as Very, and partnering with other familiar fashion retailers such as River Island and Top Shop. Like many retailers, SD have a range of discounts to perpetuate and strengthen the relationships and ultimately purchases from customers, such as vouchers issued for current purchases and general onsite promotions. SD were interesting in understanding who the customers are who are using discounts; there was a prevailing view in SD that discounts were not adding value to the company as the customers who were taking advantage of them the most were low-value customers.

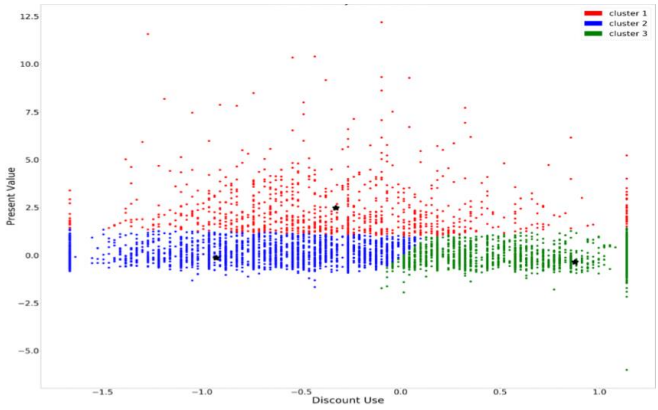
Data and Methods

SD hold their data in Teradata and is accessed via SQL. Discount use amongst customers was dispersed as seen from the graph below:



Clustering seemed like an obvious place to start, due to the fact that it is a conventional ay of understanding customer segmentation. This was performed using K Means clustering, with a different number of features tested. The Elbow Method was used to identify the optimal number of clusters, and the key variable of interest – discount use, was clustered with value, number of orders and a number of other different variables. A Hidden Markov Model was considered to see if it could be identified why customers move from a state of 'low discount use' to 'mid' or 'high discount use', however, the scenario under question did not meet a key Markov assumption (the Markov Chain). Therefore, a logistic regression was used instead to identify which features increased (and conversely which features decreased) the likelihood of customers belong to a particular discount usage category.

Key Findings



High value customers are not discount dependent (high discount users) nor discount averse (low discount users): cluster analysis with a different number of variables found high value customers to be mid-discount users (cluster group 1 in the diagram above).

This is expected, as it is irrational to receive a voucher and not use it.

Furthermore, the highest discount users are also the group with the lowest value (cluster 3). This finding is maintained when clustering on either two or three variables. The difference between high discount customers and low discount customers is relatively small in terms of present value and shopping frequency. Different features are important when predicting different types of discount. However, value, average number of days between orders and number of orders usually feature.

Value of the Research

This research enabled SD to understand the profile of customers in relation to discount usage, based on data. This has implications for retailers who use discounts; understanding how discounts are being used can inform whether or not to maintain a discount strategy based on the value that it adds to the company in question.