

Customer segmentation using spatio-temporal data

Ellen Talbot¹, Alex Singleton¹ and Dean Riddlesden²

¹University of Liverpool, ²Boots

Project Background

Shopping habits are changing; when, where and how. With data from a major High Street Retailer's loyalty card members, this research project aims to begin to understand the when. Literature already covers much of the where and how; the digital revolution has reshaped the landscape of the high street immeasurably, and the literature is well informed on how people are using the freedom the internet offers in order to shop 24 hours a day, but relatively little is known about when people visit stores, what influences them to shop and spend when they do, what for and how frequently. With a focus on big data and exploratory methods, this project looks for patterns without basing them wholly on existing theory, instead inferring possible explanations from the results themselves.

Data and Methods

The report finds and describes four distinct clusters of customers from a dataset of members of a High Street Retailer's loyalty card scheme, all of which have temporal spending differences, which could be applied in the business to create a new customer segmentation. Using the CLARA clustering algorithm overcomes the complexities of traditional data analysis and advocates the use of a data driven 4th paradigm, coping well with an extremely large set of 'Big Data'. It successfully handled a dataset of over 150 million rows by operates by considering subsets of fixed size, sampling over the entire dataset so that time and storage requirements become linear in n rather than quadratic. Extracted from each cluster were a set of values which offered insight into the temporal patterns. By creating a ranking tables from the computed results and joining this with the figures creates over the three main temporal grains; daily, weekly and monthly; it was possible to write short descriptive statements analysing the characteristics of each cluster.

Key Findings

The Pen Profiles for each cluster consolidate the analysis into short and descriptive statements and were given the following 'short and interesting' headings; Big Budget, Big Shop, those who spend the most money the most regularly; Weekday Browsers, those who are retired, browsing and spending very little;

Pocket Money Pick-ups, young people who buy cheap items fairly frequently; and Sun and Santa Shoppers, who show an increase in spending around seasonal events such as summer holidays and Christmas.

Cluster	% of observations	Mean age	Mean spend	% of females	% of males	% of Transactions AM	Mean spend - AM
1	35	44	£25.27	36	30	31	£22.84
2	30	67	£9.00	30	34	40	£8.65
3	32	29	£7.01	31	34	27	£6.11
4	3	46	£14.84	3	2	2	£13.35

Table 1. Cluster summary table - proportions

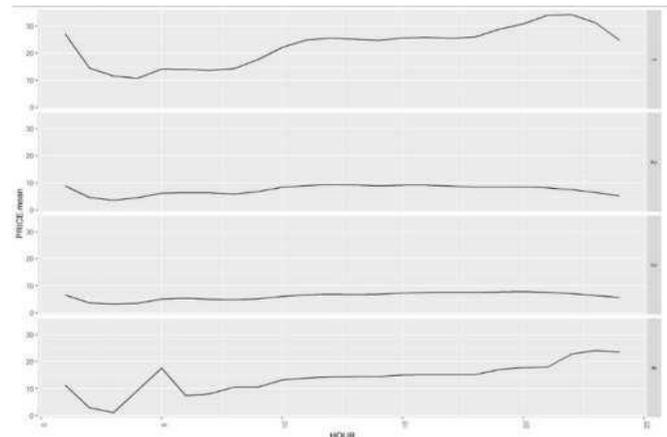


Figure 1. Average hourly spending pattern by cluster

Value of the Research

Overall this report finds that from the inclusion of temporal data into an analysis for new customer segmentations useful new patterns do emerge, which can be placed within the existing literature – for example; young people tend not to spend a lot in the High Street Retailer at Christmas because they are choosing to move their shopping online and older people are more likely to shop during the week because they are not constrained by work commitments. This research has opened the door for the continuation of research into temporal demographics and the scope for further study encompasses ideas such as the inclusion of product categories and store types to deepen the understanding of temporal patterns, for example, whether or not each cluster shops at a different store type given their needs and available time budget.