

The Dynamics of Consumption Space within Cities

Ellen Talbot, Alex Singleton, Dani Arribas-Bel¹

¹Department of Geography and Planning, University of Liverpool

November 23, 2016

Summary

The dynamics of consumption space within our cities are increasingly visible through those data collected about the location, transaction types and the characteristics of shoppers. However, such complexity is difficult to understand through traditional methods given these data typically ensue characteristics of “big data”, yet offers potential new insight about the dynamics of our cities. This study explores the temporal dynamics of how people use consumption space within our cities. This results in new patterns emerging beyond those that might be identified from attribute or locational data alone. Multiscalar temporalities showed general differences between age and gender as well as highlighting four distinct clusters and their unique consumption trends.

KEYWORDS: Geodemographics, Consumer trends, Temporal classification, Big data, Machine Learning.

1. Overview and Literature

This research project focuses on the temporal dimension of consumer behaviour which is significant to understanding social and economic geographies (Miller, 2008; Shaw and Yu, 2009). Geodemographics provides a well-established framework for understanding aggregate spatial behaviour (Longley, 2012; Singleton and Spielman, 2014), however, to date has had limited attention in the examination of time in an integrated way. Previous work has either mapped changes in classifications between censuses (Singleton et al, 2016) or explored different temporal cross sections such as workday geography (Martin et al, 2015).

2. Methodology and Data

The data used for this study comprise consumption records for an anonymous High Street Retailer. These data pertain to the dates covering the financial year 2013 and contain individual details, alongside their basket level spend through this period of time. In this study, the retailer is referred to as “High Street Retailer”, and were made available through the ESRC Consumer Data Research Centre. Given the retailer, the data are skewed towards females of child bearing age. The National Consumer Agency (2013) presented research which suggests that 72% of women take responsibility for the households shopping, and the loyalty card is more likely to be in their name despite making purchases for the whole family.

¹ e.talbot@liverpool.ac.uk
alex.singleton@liverpool.ac.uk
d.arribas-bel@liverpool.ac.uk

2.1 Cleaning and enriching the data

The first stage in the analysis was to understand the characteristics of the data. The entirety of the analysis was completed within the R language inside a secure data centre. The data was aggregated from product to ‘basket’ level to enable analysis of a total spend, and then cleaned to remove outliers, null values and those with an invalid address, as described in **Figure 1**. This extra step was implemented following a report from the High-Street Retailer suggesting this is an element of data cleaning that they were interested in, but served added purpose in this report to ensure that the clustering results were not skewed by people appearing to travel extreme distances to their local store, when they have simply moved away.

Table 1 below shows the cumulative effects and steps of the data cleaning which took place before the analysis took place.

Table 1 The cumulative effects of data cleaning

<i>Step 1: Pre data cleaning</i>
Total number of records: 463,282,609
<i>Step 2: Aggregated 'basket' level transactions</i>
Total number of records remaining: 220,491,395
Average number of products per basket: 2.1
<i>Step 3: Removing account holders who have an invalid address</i>
Number of baskets removed: 58,358,668 (26.5%)
Total number of baskets remaining: 162,132,727
<i>Step 4: Removing records which contain invalid information</i>
Number of baskets removed: 4,314,454 (2.6%)
Total number of baskets remaining: 157,818,273
Total number of unique account holders: 14,154,512
Average number of baskets per account holder: 11.1
Final number of 'baskets' to be included in the analysis: 157,818,273

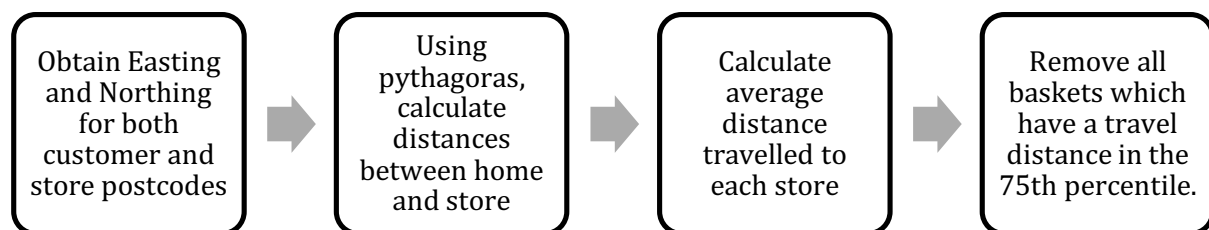


Figure 1 Invalid account addresses

2.1 Descriptive statistics

The first step in the analysis was to understand the general trends in consumption at a number of different temporal granularities. The date and time stamps were transformed so it was possible to distinguish AM, PM, lunchtime and evening trends, weekday and weekend trends and seasonal

changes. Age and Gender were investigated at all multiscalar temporalities to gain an overview of general trends. Once the clustering was complete, similar descriptive outputs were produced to expand descriptions that were created in the pen portraits.

3. Mapping Dynamic Consumption Spaces

Comparison between consumer characteristics was achieved using the CLARA clustering algorithm, which overcomes the complexities of traditional data analysis and was chosen instead of using a standard partition around medoids (PAM) algorithm because of its capacity for handling extremely large datasets without the need for enormous computing power. It successfully handled a dataset of over 150 million rows by considering subsets of fixed size, sampling over the entire dataset so that time and storage requirements become linear in n rather than quadratic.

Two instances of CLARA were run and the resulting outputs analysed to decide the optimum number of clusters – smaller dissimilarities are preferable; therefore, four clusters were calculated. Extracted from each cluster into a summary table were a set of values which offered insight into the temporal patterns. A framework based on previous work done by Singleton and Spielman (2015) was implemented to aid the classification and analysis of the clusters, covering membership, temporality and spending, defining which variables were included in the final analysis. A table of raw data was extracted and used to aid the creation of ranking labels. Joining this with the figures created over the three main temporal grains; daily, weekly and monthly; it was possible to write short descriptive statements analysing the characteristics of each cluster and turn these into a ‘pen profile’.

4. Results and Conclusions

The descriptive statistics showed differences in gender and age spending. It was found that middle aged people made up the largest proportion of baskets and accounted for the highest spending, most likely because they are of child bearing age and are likely to have a young family with additional requirements. Those who did not disclose their gender were found to spend the most, however it would require more evidence possibly in the form of market research or interviews with customers to get a fuller understanding of why this may be the case. Those at the youngest or oldest end of the age range were the lowest spenders, perhaps related to their lower levels of disposable income, but also perhaps because they do not have a requirement for family health products, in which the HSR specialises.

The Pen Profiles for each cluster consolidate the analysis into short and descriptive statements and were given the following names: 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.

In summary, the inclusion of temporal data into the analysis of customer segmentations highlights useful new patterns, which can be situated 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 each cluster shops at a different store type given their needs and available time budget.

5. Acknowledgements

With thanks to the Consumer Data Research Centre and the High-Street Retailer for supporting the

CDRC sponsored dissertation.

6. Biography

Ellen Talbot is a First Year PhD Student at the University of Liverpool's Geographic Data Science Lab and joined the team in 2015 undertaking the Geographic Data Science MSc. Research interests include new temporal flows and geodemographic classifications composed from consumer and energy datasets.

Alex Singleton is a Professor of Geographic Information Science at the University of Liverpool and Deputy Director of the ESRC Consumer Data Research Centre. His general research is concerned with how the complexities of individual behaviours manifest spatially and can be represented and understood through a framework of geographic data science.

Dani Arribas-Bel is a Lecturer in Geographic Data Science at the Department of Geography and Planning, and a member of the Geographic Data Science Lab at the University of Liverpool where he also directs the MSc in Geographic Data Science. His main research interests are urban economics and regional science, spatial analysis and spatial econometrics and open source scientific computing.

References

Fletcher, G., Greenhill, A., Griffiths, M. and McLean, R (2016) "The social supply chain and the future high street" *Supply Chain Management an International Journal* Vol 21 (1) pp 78 – 91.

Longley, P. (2012) Geodemographics and the practices of geographic information science. *International Journal of Geographic Information Science* Vol 26 (12) pp2227 – 2237.

Martin D., Cockings, S. and Leung, S. (2015) Developing a flexible framework for spatiotemporal population modelling. *Annals of the Association of American Geographers* Vol 105 (4) pp 754-772

National Consumer Agency (2013) Consumer Shopping Behaviour August 2013. Available online at [<http://www.cpc.ie/consumer-shopping-behaviour-august-2013>] accessed on [27.11.16]

Singleton, A. and Spielman, S. (2014) The Past, the present and the future of Geodemographic Research in the United State and United Kingdom. *The Professional Geographer*. Vol 66(4) pp.558 – 567

Singleton, A., Pavlis, M. and Longley, P. (2016) The stability of geodemographic cluster assignments over an intercensal period. *Journal of Geographical Systems* Vol 18, (2) pp 97-123

Spielman, S.E. and Singleton, A. (2015) Studying neighborhoods using uncertain data from the American community survey: a contextual approach. *Annals of the Association of American Geographers*, 105(5), pp.1003-1025.