Journal of Retailing and Consumer Services 30 (2016) 271–278

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser

Deriving age and gender from forenames for consumer analytics

Guy Lansley, Paul Longley *

Department of Geography, University College London, Gower Street, London WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 8 September 2015
Received in revised form
5 February 2016
Accepted 10 February 2016

Keywords:
Age
Gender
Geodemographics
Big Data
Social media

ABSTRACT

This paper explores the age and gender distributions of the bearers of British forenames and identifies key trends in British naming conventions. Age and gender characteristics are known to greatly influence consumption behaviour, and so extracting and using names to indicate these characteristics from consumer datasets is of clear value to the retail and marketing industries. Data representing over 17 million individuals sourced from birth certificates and market data have been modelled to estimate the total age and gender distributions of 32,000 unique forenames in Britain. When aggregated into five year age bands for each gender, the data reveal distinctive age profiles for different names, which are largely a product of the rise and decline in popularity of different baby names over the past 90 years. The names database produced can be used to infer the expected age and gender structures of many consumer datasets, as well as to anticipate key characteristics of consumers at the level of the individual.

1. Introduction and overview

The advent of new sources of consumer data, such as those arising from the use of social media, online shopping and customer loyalty databases, present new opportunities to measure and model the activity patterns of individuals. Such data may be related to detailed functional taxonomies of the locations that consumers visit or the products that they buy, bringing insight into the nature and likely motivations for observed activity patterns. But there has been no commensurate improvement in the detail with which we are able to characterise the individuals themselves, and thus ascertain how representative they are of consumer segments or indeed the population at large.

In this context, our own attempts to understand consumer behaviour have become focused upon the task of front-loading the inferences that may be drawn from consumer names and social media user identifiers, in order to relate new Big Data sources to the wider populations from which they are drawn. In this paper we describe some of the ways in which individual given (fore-) names can be analysed in order to ascribe age and gender characteristics, as part of this task. The work builds upon a commercial classification (Monica: CACI, London) and is part of a wider research programme which explores given and family name pairings in order to infer individual characteristics in consumer research (Mateos et al., 2011; Longley and Adnan, 2016).

A person's given name can be used to infer a number of key individual characteristics, as a result of the ways in which names are typically distributed according to age, gender and ethnicity. Most forenames are gender specific and many can be traced to ethnic groups through common heritage (Mateos et al., 2011). There has been much research into patterns in names and how they link to cultural heritage and wider society. However, only limited attempts have been made to consider the age and gender distributions of forenames in Britain (Finch, 2008). Names vary by gender and age across time largely because of shifts in popularity of different baby names over time and the influence of migration. This paper summarises work undertaken to model the age and gender structures of different name holders in the Great Britain. From achieving a greater understanding about the demographic characteristics of different name bearers, more information can be inferred from consumer data which include names but no further details about the individuals.

Retailers have benefitted greatly from geodemographic datasets, made available from the government or other businesses, as a means of segmenting and understanding consumers. Such data allow retailers to plan their stock and marketing accordingly to the local population characteristics (Mitchell and McGoldrick, 1994; O'Malley et al., 1997). However, this conventional approach is based upon the assumed correspondence between night-time residence and consumer behaviour (Harris et al., 2005). The data is also typically aggregated and/or modelled, therefore the traits for each spatial unit may not be entirely representative of every resident (Openshaw, 1984). Consequently, there have also been developments in micro-economic modelling which seek to place the

* Corresponding author.
E-mail addresses: g.lansley@ucl.ac.uk (G. Lansley), p.longley@ucl.ac.uk (P. Longley).

http://dx.doi.org/10.1016/j.jretconser.2016.02.007

focus upon the individual to arguably provide an inherently superior approach to understanding consumer behaviour as it circumvents any issues of ecological fallacy (Hensher and Johnson, 1981; see also Longley and Adnan (2016)). However, in order to apply such techniques individual level data is required, and often consumer data is absent of demographic information. For instance, the records of an online account associated with large retailer will typically only include a name, address and purchase history. Therefore, inferring demographic traits from names data could allow analysts to harness more consumer insight from many data sources.

1.1. Demographics and consumption

In this paper we take it as axiomatic that gender and age of consumers both heavily influence behaviour and consumption practices. Consequently, much research has been devoted to understanding the influence of demographics on consumption, such as the incorporation of such characteristics in product and brand choice models (Kalyanam and Putler, 1997), and identifying target demographics for new products. Geodemographic segmentations are widely used to segment the population in to distinctive consumer groups from multivariate data (Harris et al., 2005), and age and gender are included in them as key correlates of consumer behaviour.

The gender divide in consumption practices is most obvious amongst certain product types such as clothes and cosmetics, where retailers produce and sell entirely different lines of stock tailored for each sex (Scanlon, 2000). Consumption has become an important means through which individuals construct their gendered identities (Baudrillard, 1998). There are even gendered variations in consumption of products which are not produced exclusively for one gender. For instance, women have been found to be more health conscious as exerted by their food consumption practices and therefore they are generally likely to perceive certain foods differently to men (Wardle et al., 2004). There is even a gender divide in perceptions of shopping behaviour, women being traditionally more likely to perceive shopping as a leisure experience and therefore spend more time visiting high streets (Campbell, 1997; Lunt and Livingstone, 1992). By contrast, males have been traditionally the most dominant patrons of online shopping websites (Dittmar et al., 2004; Rodgers and Harris, 2003), and are more comfortable using multiple channels when making a purchase (Blázquez, 2014).

Age is also an influential characteristic of consumer behaviour, in terms of product and brand preference, and also in terms of how individuals shop. For instance, younger consumers are usually more likely to patronise online shopping channels, whilst older individuals are typically less engaged with the Internet and other modern shopping channels such as mobile commerce (Sorce et al., 2005). Consumer behaviour is greatly influenced by the family life cycle and the ways in which disposable incomes are channelled through consumption (Reynolds and Wells, 1977). Amongst the adult population, different cohorts are known to have different consumption practices which link to their life stage, their physical and health characteristics, and their cultural characteristics. Shared experiences during adolescence and beyond that are traceable to societal, cultural and environmental traits can encourage individuals to develop values that they will retain over time and can give rise to “cohort effects” (Harmon et al., 1999). Consequently, those from an age cohort may share distinctive values and this is likely to influence their consumer behaviour (Pentecost and Lynda, 2010).

Understanding the demographic characteristics of consumers is therefore very important to developing a sustainable retail strategy. Consequently, inferring demographic traits of consumers has been a vital area of marketing and consumer research (McDonald, 1995; Carpenter and Moore, 2006). It is also important to adapt to local and national demographic changes. For instance, there is an increasing imperative to understand the consumption practices of elderly consumers in many western countries given their aging populations (Kohijoki and Marjanen, 2012). Even amongst stable populations, previous research has established that age cohorts have unique consumption traits relative to previous generations (Bakewell and Mitchell, 2003). Although every individual may have distinctive tastes, general consumption practices nevertheless vary by age and gender. Therefore, the possibility of estimating the general demographic structure alongside a consumer's distinctive personal characteristics from individual customer records can be a fruitful means of obtaining key information about clients and customers.

2. Names and demographics

There is a wide range of Big Data sources on the population which includes name identifiers, but have little or no additional demographic information. These include electoral registers, customer records and social media data. There have subsequently been attempts to harness information from names by examining how names are distributed through contemporary society. Perhaps the most sophisticated developments have been in the production of cultural, ethnic and language group classifications from forename–surname pairs (Mateos, 2007; Mateos et al., 2011). Surnames, in particular, can identify bonds between family members, and therefore can be aggregated to represent distinctive cultural groups.

Historically, there has been a range of processes that influence popular naming conventions in the Great Britain (Smith-Bannister, 1997). Over the last century, popular naming practices have become far more erratic (Galbi, 2002), reflecting secularisation of society, migration trends and social mobility. The UK Office for National Statistics (ONS) has nevertheless identified clear trends in baby naming over the years (Matheson and Summerfield, 2000), and names therefore offer a viable means of estimating age structures from larger populations (Scharf, 2005).

We associate names with their bearers, yet forenames more directly manifest the predilections, priorities and preferences of either or both parents (Gureckis and Goldstone, 2009). The choice of baby name is likely to vary systematically between parents from different socio-economic backgrounds and cultural groups. The favourability of names is also influenced by popular trends, giving rise to temporal autocorrelation in name frequencies (Xi et al., 2014). Research has identified that parents in the USA perceive baby names which are growing in popularity to be more desirable than those whose popularity is waning (Berger and Mens, 2009; Gureckis and Goldstone, 2009). In additional to shifts in societal values, more subtle environmental and internal influences also drive the popularity of names for particular groups (Lieberson, 2000). Whilst the choice of baby name is influenced by various sociological factors, some names are handed down by family members or have remained popular because of links to cultural heritage: such names are much less likely to vary much between age groups (Finch, 2008).

Forenames are subsequently an important part of an individual's identity and can even act as a positive source for cultural capital (Lord, 2002). Observers may associate names with stereotypes, such as a child's likely educational attainment (Harari and McDavid, 1973; Erwin and Calev, 1984). This may be grounded in truth because of the different forename preferences of parents who themselves have different experience of, and attitudes to, educational capital formation.