

# **Postcode Based Factor Scores Using the 2011 Census Data**

## **1. Introduction**

Using data from a variety of sources we have created a series of highly predictive factor scores which can be used in a wide variety of marketing applications such as predictive targeting, profiling and modelling, catchment area analysis for store or site location.

## **2. Information Used to Create the Factor Scores**

### **2.1 2011 Census**

The 2011 Census collected information from individuals and households on a wide variety of topics. For individuals these include

- sex
- age
- whether in full time education
- ethnic group
- religion
- limiting long term illness
- provision of unpaid care
- mobility
- educational and professional qualifications
- employment status
- address of employment and business of employer
- method of travel to work
- hours worked

and for households these include

- number residents in households
- type of accommodation
- number of rooms
- provision of amenities
- cars and vans available to the household

- tenure
- relationships between members of a household

This is a fairly comprehensive list, and provides a detailed picture about the population and how they live. However, in addition, a number of derived classifications are also provided by the Office of National Statistics. These include:-

- family composition (eg single person pensioner, married couple with one child etc.)
- social grade (a classification widely used in market research)
- ONS standard occupational classification
- National Statistics Socio Economic classification (NS-SEC)
- Standard industrial classification (SIC) which classifies businesses by type of economic activity

All of this information is available for a range of geographic areas but the most detailed level is the output area. In England and Wales the average size of an output area is around 125 households whilst for Scotland the average size is around 50 households. The output area data was used to derive the factor scores.

## **2.2 Other data Sources**

In addition to Census data we also used output area Council Tax information in the derivation of the factor scores.

## **3. Factor Analysis Results**

In order to make all of this data more usable we have simplified the data using a technique called factor analysis. Factor analysis is a statistical technique used to identify a relatively small number of factors that can be used to represent relationships amongst larger sets of interrelated variables. In essence the technique simplifies the data and identifies underlying dimensions or factors within the data. These dimensions are held as a series of statistically independent scores that represent the underlying dimensions. The factor scores are:

### **Factor 1 - Professionals and Affluence**

Variables most positively associated with this group are

- Higher managerial & professional occupations
- High levels of educational qualifications
- Social grade A/B
- Households with 7 or more rooms
- Expensive properties in council tax bands G,H, or I

Variables most negatively associated with this group are

- Social Grade D or E
- Elementary (unskilled) occupations
- No qualifications
- Sales and customer service occupations
- Social renting

### **Factor 2 – Mortgagees**

Variables most positively associated with group are

- Owned with a mortgage
- 3 or 4 people in household
- 2 or more cars
- Modest levels of educational qualifications
- Above average numbers of economically active persons per household

Variables most negatively associated with this group are

- Living in purpose built or converted flats

### **Factor 3 – Young Families**

Variables most positively associated with group are

- Aged 25-44
- Pre-School aged children

- In full time employment
- Private Renting

Variables most negatively associated with group are

- Aged over 55
- Owns home outright

#### **Factor 4 – Students**

Variables most positively associated with group are

- Aged 18-24
- Full Time Students
- Level 3 Qualifications (entry level qualifications for higher education)

#### **Factor 5 – Large Families**

Variables most positively associated with group are

- 5 or more people in the household
- People aged 5-11
- People aged 12-17

Variables most negatively associated with group are

- 2 person households

#### **Factor 6 – Self Employment**

Variables most positively associated with group are

- Small Employers and Own Account Workers
- Skilled Trades
- works from home

### **Factor 7 – Semi-Detached Suburbia**

Variables most positively associated with group are

- Properties in council tax band C
- Semi-Detached Properties

Variables most negatively associated with group are

- Properties in council tax band A

### **Factor 8 – Terraced Properties**

Variables most positively associated with group are

- Terraced Properties
- 1 car households

### **Factor 9 – White Collar**

Variables most positively associated with group are

- Administrative and Secretarial Occupations
- Social Grade C1

### **Factor 10 – Middle England**

Variables most positively associated with group are

- Properties in council tax band D

#### **4. The Data Held**

The factor scores were derived for each census output area and then linked to postcodes so that all postcodes within the same output area has the same set of scores.

Each factor score is held in two formats. The first is a standardised numeric score with a mean of zero and standard deviation of one. Thus a large positive score denotes the presence of people or households with the attribute in question. Whilst a large negative score denotes that very few people or households with that attribute are present. The numeric score tells you exactly how far up or down the scale any given postcode falls. An affluence score of +1.96 or more, for example, would tell you that this postcode falls within the top 2.5% of the country in terms of its affluence rating. These scores can be used directly in analysis and statistical modelling. The second format in which the information is held is as a banded ranking. Each postcode factor score is grouped into one of twenty possible bands so that 5% of all households fall within each band. Thus a band one ranking for affluence denotes that the postcode in question contains some of the most affluent households in the country and is within the top 5% in the country, whilst a band 20 ranking denotes that the postcode contains some of the least affluent households. Using the banded ranking it is possible to overlay the banded scores onto any postcoded source file and generate a profile of that file by simply comparing the proportion of the file in each band against the expectation. If the band proportion is significantly in excess of 5% this shows where the file is over represented and contains more people or households than you would expect by chance, whereas bands with significantly less than the expected 5% show an under representation. This makes it possible to profile any name and address based file such as a group of customers or a group of good customers, or it could be potential customers or customers who have purchased a particular product or even lapsed or ex-customers.

**David Griffiths**

**18<sup>th</sup> January 2015**

**TRAC Consultancy**

**Email** [daveowengriffiths1@gmail.com](mailto:daveowengriffiths1@gmail.com)

**Website** [www.tracconsultancy.co.uk](http://www.tracconsultancy.co.uk)