

# Postcode Based Factor Scores Using the 2001 Census and Other Data

## 1. Introduction

Information is the key to marketing success. The more information you have about people the more successful you are likely to be in finding and retaining customers. 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 2001 Census

The 2001 Census was held on 29 April 2001 and 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 such as baths/showers, wc's and central heating
- 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 to this data 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 linked to postcodes and used to derive the factor scores.

## **2.2 Other data Sources**

In addition to Census data we also used the Research and Analysis Consultancies predictors of wealth and consumer activity in the factoring. These are two postcode based predictors which describe levels of wealth and consumer activity for each postcode in the country.

## **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 large sets of interrelated variables. In essence the technique simplifies the data and identifies underlying dimensions or factors within the data.

The results of factoring are a series of statistically independent scores that represent the underlying dimensions within the data. In this case the results are:

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

Variables most positively associated with this group are

- High levels of educational qualifications
- SEC Group Higher Managerial
- Social Grade A/B
- The wealth score
- Households with 7+ rooms

Variables most negatively associated with this group are

- SEC Group Semi-Routine occupations
- Social Grade C2
- SEC Group Lower Supervisory
- SEC Group Routine occupations
- Social Grade D
- Low Levels of Educational Qualifications

### **Factor 2 – Families with School Aged Children**

Variables most positively associated with group are

- Average number of children per household
- Families with 2 or more children
- People aged 5-11
- People aged 12-17
- 5 people in household
- High levels of consumer activity, particularly mail order

### **Factor 3 – Mortgagees**

Variables most positively associated with group are

- SEC Occupational Group, Intermediate
- Buying home on a Mortgage
- High levels of consumer activity generally
- Social Grade C1
- Full Time Employee
- Educational Qualifications Level 2 (middle order)
- 2 Dependent Children in household
- 2 cars

Variables most negatively associated with group are

- Rents Home
- SEC Occupational Group, Never Worked
- Unemployed
- Lone Parents

### **Factor 4 – Younger Pensioners and Empty Nesters**

Variables most positively associated with group are

- Household Composition Family Pensioners
- Aged 65-74
- Owns Home Outright
- Aged 55-64
- Occupancy Rating +2 (2 or more rooms compared to people)

Variables most negatively associated with group are

- Aged 25-34
- Single Person Households (Not Pensioners)

**Factor 5 – Self Employment**

Variables most positively associated with group are

- Small Employers
- Self Employed with employees
- Self Employed without employees
- Hours Worked 49 or more
- Second Home
- Home working

**Factor 6 – Students**

Variables most positively associated with group are

- Household Composition All students
- Aged 18-24
- Full time students
- Educational Qualifications Level 3 (pre degree)
- Tenure Private Rented

**Factor 7 – Young Families**

Variables most positively associated with group are

- Families with 1 child 0-4
- Aged 0-4
- Household Composition – Couples 1 child

**Factor 8 – Detached Properties**

Variables most positively associated with group are

- Cars in household 3
- Cars in household 4
- Detached Property
- Company Directors

Variables most negatively associated with group

- Cars in household 1

**Factor 9 – Asian Families**

Variables most positively associated with group are

- Ethnic Group Asian
- No of People in household 6 or more
- Household Composition Other with children

**Factor 10 – Couples**

Variables most positively associated with group are

- Number of people in household 2
- Household Composition Couples no Kids
- Families with no dependent children

Variables most negatively associated with group are

- Single Person Pensioner Households

### **Factor 11 – Older Families**

Variables most positively associated with group are

- Families with 1 child 12-18
- Number of people in household 3
- Couples with non-dependent children
- Aged 45-54

### **Factor 12 – Overcrowded Accommodation**

Variables most positively associated with group are

- Ethnic group white

Variables most negatively associated with group are

- Occupancy Rating -2 ( 2 or more people than rooms)
- Ethnic group Chinese
- Ethnic group Black

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 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 expected 5%. If the band proportion is significantly in excess of 5% this shows where the file is over represented and contains more people than you would expect by chance, whereas bands with significantly less than the expected 5% show where the file is under represented. The file in question could be more or less any group such as a group of customers or better still good customers, or it could be potential customers or customers who have purchased a particular product or even lapsed or ex-customers.

**David Griffiths**  
**TRAC Consultancy**  
**18.10.2003**