



**The Use of Census & Other Data in the
Build of Geo-Demographic Segmentation
Systems for Identifying Consumer
Behaviour**

David Griffiths

Contents

- **The Use of 2001 Census Data**
- **The integration of Census with other data**
- **The use of different statistical techniques and the building of different cluster solutions**
- **Evaluating which solution is best – Sonar**
- **Looking at how Sonar performs in the real world**
- **Conclusion**

2001 Census Data

•Published for variety of Geographic Areas. The most detailed is Output Areas. Approximately 220,000 in GB

Person Level:

Gender

Age

Whether in full time education

Ethnic group

Religion

Limiting long term illness

Provision of unpaid care

Mobility

Educational & Professional qualifications

Travel to work

Hours worked

Household Level:

Number of residents in household

Type of accommodation

Number of Rooms

Provision of amenities

Car ownership

Tenure

Household composition

SIC- Classification

Social Grade, SEC

Additional Data Used in Sonar

- **Postcode Measures of Wealth**
- **Postcode Measures of Consumer Activity**
- **Land Registry 2002 House Price Data**

Additional Data Used in Sonar

- **Wealth and Consumer Activity are postcode scores which rank each postcode - from highest to lowest.**
- **Wealth score was built from a variety of data sources and includes measures about earnings, number of rooms, the presence of detached properties, house prices, and occupational and unemployment data.**
- **Consumer activity ranking comes from a variety of data sources and includes measures on product holdings, levels of residential property sales, the level of home ownership and mortgaged properties, new car ownership and insurance ratings.**

Geo-Demographics

- **Huge growth in use of Geo-Demographics in last three Decades:**
 - **Easy to understand and use.**
 - **Present a picture of local Geographies which in most cases is plausible and insightful.**
 - **Proven discriminatory power.**
 - **Neighbourhoods differ in purchasing behaviour and attitudes in predictable and significant ways.**
- **Predicated upon the 'birds of a feather' principle.**
 - **People who live in the same neighbourhood are more likely to be alike than you would expect by chance.**

TRAC/The Clockworks - aim was to build the best Geo-Demographic System - Sonar

- **Tested 3 Statistical Techniques:**
 - **Two Stage Clustering**
 - **Kohonen Neural Network**
 - **McQueens K Means**
- **Tested different combinations of variables in clustering**
- **Tested different numbers of clusters in the solution**

In all over 100 different solutions (combinations of the above three were built)

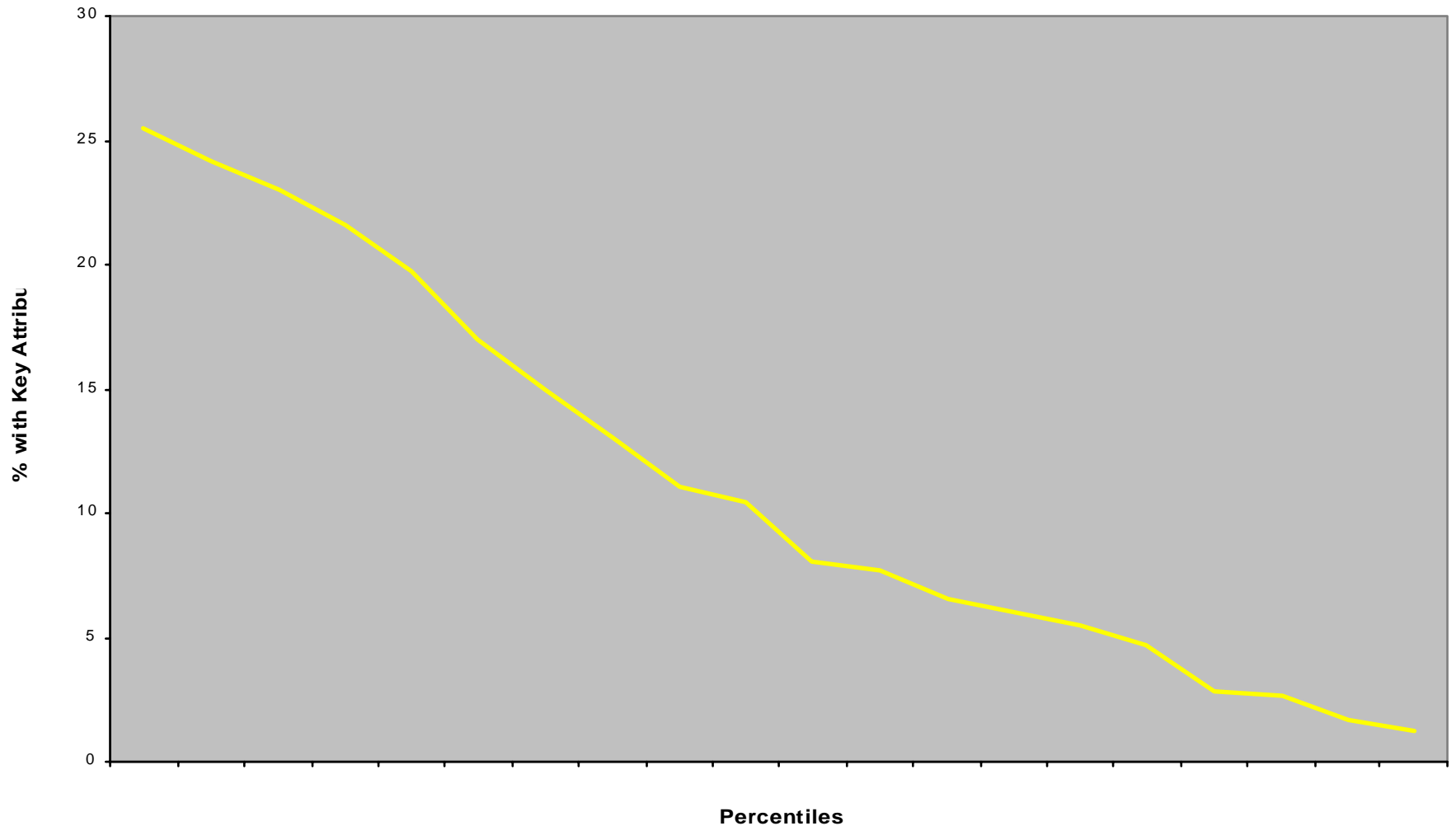
The Preferred Solution - Sonar

- **The Sonar Geo-Demographic product was the solution which outperformed all other solutions in terms of ranking closeness.**
- **No one solution performed better across all key attributes.**
- **One solution emerged which was frequently the top ranked solution. This solution was a Two Stage Clustering solution with 80 clusters - Sonar.**
- **Sonar has reasonably evenly sized clusters.**
- **Sonar performed particularly well on wealth, occupational related measures and also on measures related to household composition and family structures.**

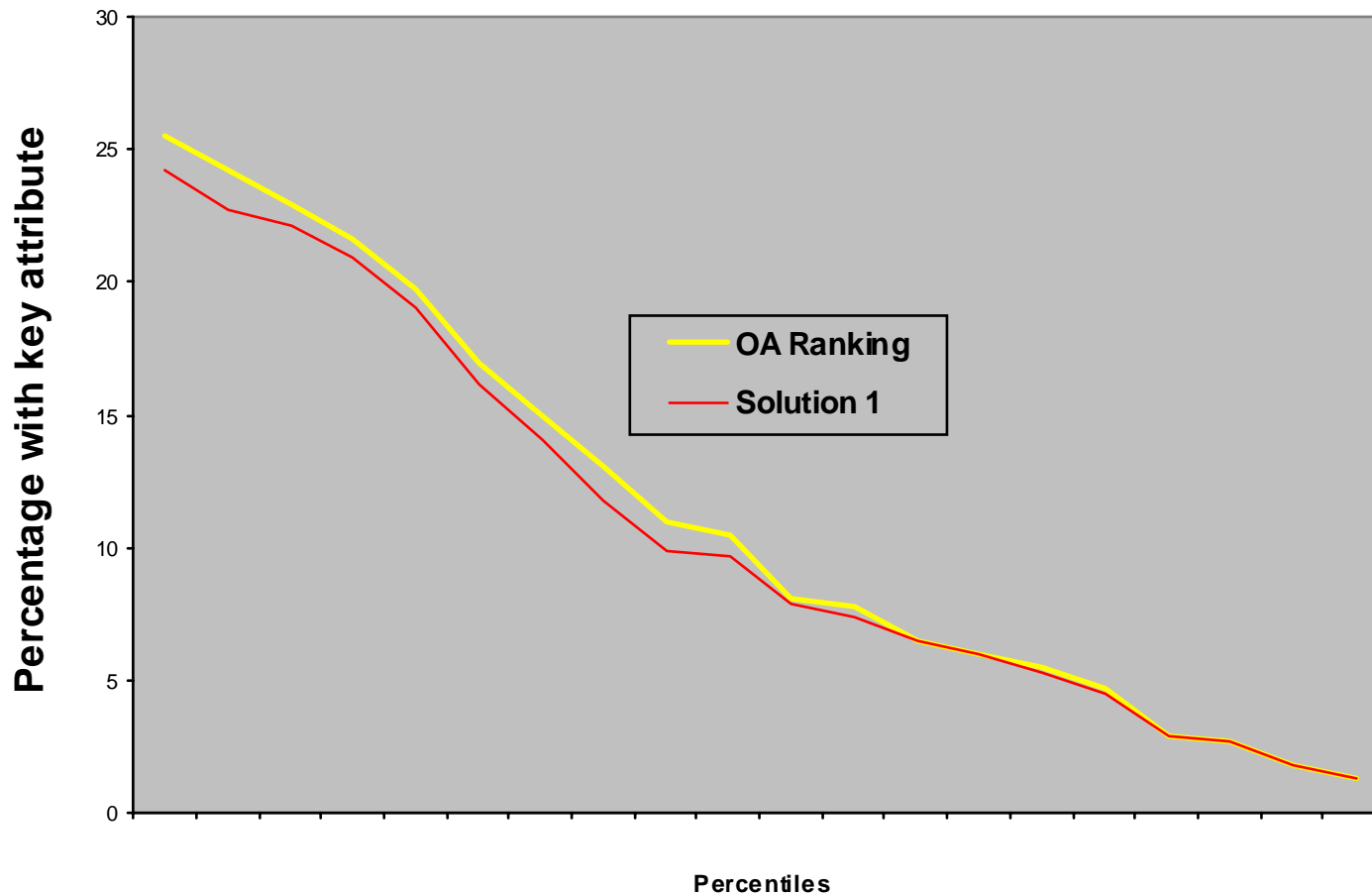
Evaluating which is the 'Best' Solution?

- **The construction of a Geo-Demographic classification for targeting has traditionally presented a problem of how to measure performance.**
- **No clear statistical function against which to measure the performance of alternative solutions.**
- **Approach adopted to this problem can be outlined by a number of commentators as follows:**

Output Area Ranking - % with a key attribute



Output Area Ranking - % with a key attribute



Cluster Solution Evaluation

- Each of the different solutions were compared across a range of key attributes which included:

Age

Household composition

Number of people in household

Occupation

Tenure

Number of rooms

Car ownership

Amenities

Two Step Clustering Overview

- **Clustering Using Two Step Cluster Analysis**

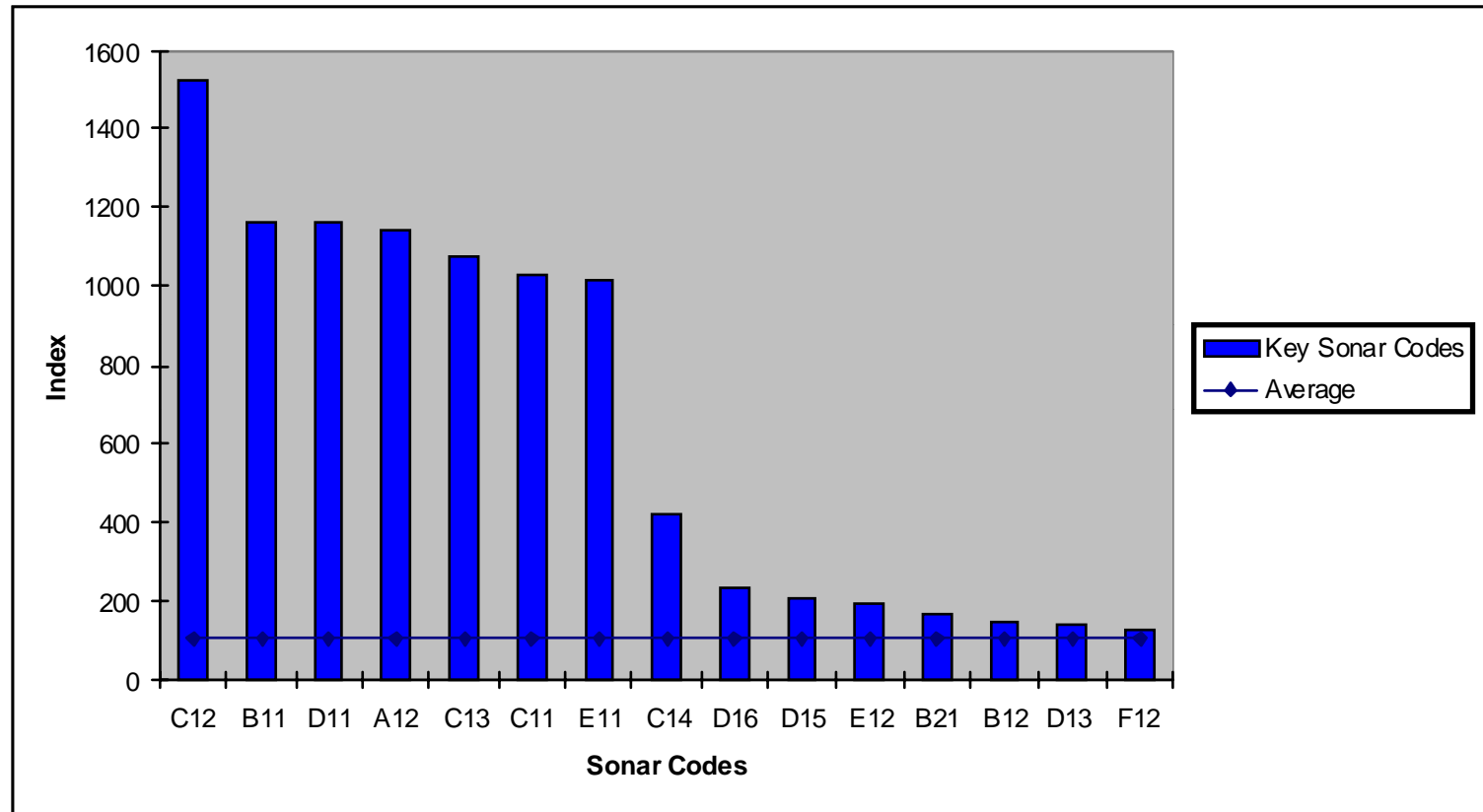
- **Step 1**

- The procedure begins with the construction of a Cluster Features Tree
 - The first case is placed in a node at the root of the Tree
 - Each successive case is then added to an existing node or forms a new node, based upon it's similarity to existing nodes using a likelihood distance measure

- **Step 2**

- The leaf nodes of the Tree are then grouped together using an agglomerative clustering algorithm. The agglomerative clustering can be used to produce a range of solutions
 - To determine which number of clusters is best each of the cluster solutions is compared using one or two possible "best fit" criteria

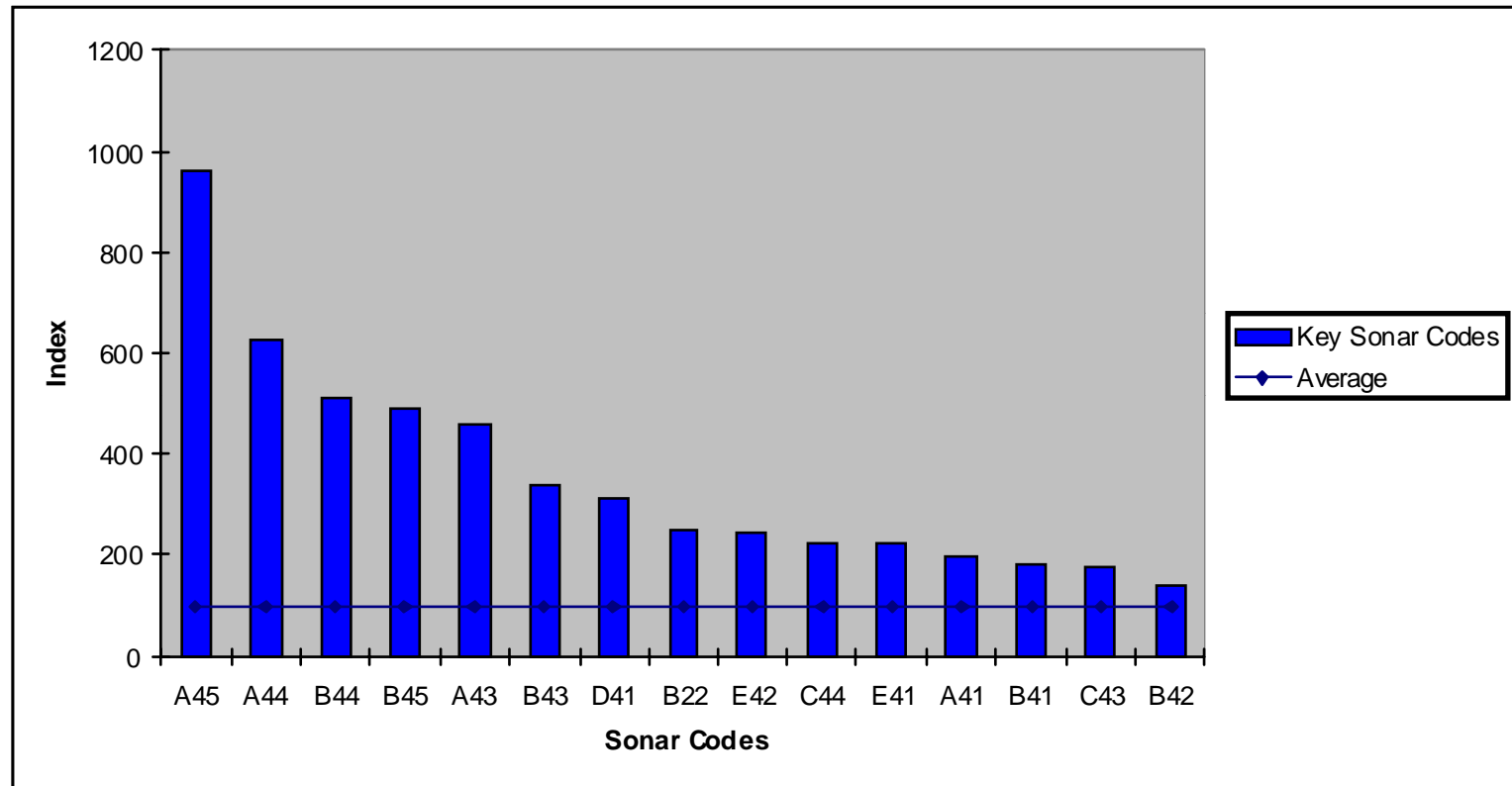
Profile of up market car buyers



Profile of up market car buyers – key Sonar codes

Sonar Code	National %	Car Buyers %	Index
C12 - Affluent Family Suburbs	0.7	10.6	1522
B11 - Affluent Young Professionals	1	11.6	1165
D11 - Professionals with Older Families	1.41	16.3	1159
A12 - Tom Brown's School Days	0.13	1.5	1145
C13 - Hardworking Entrepreneurs	0.88	9.5	1080
C11 - Affluent Young Families in Town Houses	0.46	4.7	1026
E11 - Established Professionals	0.7	7.1	1019
C14 - Commuting Life	1.36	5.7	419
D16 - Hardworking Materialists	1.77	4.1	232
D15 - Small Town Professionals	1.49	3.1	207
E12 - Bricks and Mortar Assets	1.6	3	187
B21 - Terraced Start Ups	1.17	1.9	162
B12 - Young Professionals with Pre-School Kids	1.27	1.9	150
D13 - Green Belt Expansion	1.93	2.7	140
F12 - Affluent Elders	0.85	1.1	129

Profile of down market loan take up



Profile of down market loan take up – key Sonar codes

Sonar Code	National %	Loan %	Index
A45 - Lone Parent Hardship	0.25	2.4	960
A44 - Council Sink Estates	0.56	3.5	625
B44 - Benefit Dependent Families	0.8	4.1	513
B45 - Terraced Industrial Heartlands	2.05	10.1	493
A43 - Overcrowded High Rise	1.2	5.5	458
B43 - Struggling Families	1.12	3.8	339
D41 - Council Right to Buy	0.9	2.8	311
B22 - Young Military Families	0.2	0.5	250
E42 - Hard Up & Hard Pressed	1.11	2.7	243
C44 - Post War Council Semi's	2.34	5.3	226
E41 - Run Down Provincial High Rise	0.71	1.6	225
A41 - Black Metropolis	0.81	1.6	198
B41 - Poorer Renters	1.16	2.1	181
C43 - Transitional Blue Collar	2.05	3.6	176
B42 - Asian Extended Families	0.64	0.9	141

Conclusion

- **TRAC/The Clockworks have used Census & other unique data to build a Geo-Demographic Classification called Sonar**
- **Sonar evolved from extensive testing to optimise performance**
- **The benefits of Sonar are now being realised and it offers a superior alternative to traditional systems**

s o n a r