

AGEIS MAJOR

1. What is Ageis Major?

Ageis Major is a product that allows you to predict the age of adults (people aged 18 or more) from their first name. It has been shown that name usage varies quite considerably over time and that this variation can be used to accurately predict the age of an individual. The information is supplied in the form of a directory which provides a series of probabilities (the probability that they are aged 18-24, the probability that they are aged 25-34 and so on) for each name. These represent the likelihood of someone with that name being of a certain age. All of the information is calculated as of 1.1.2009.

2. Information supplied

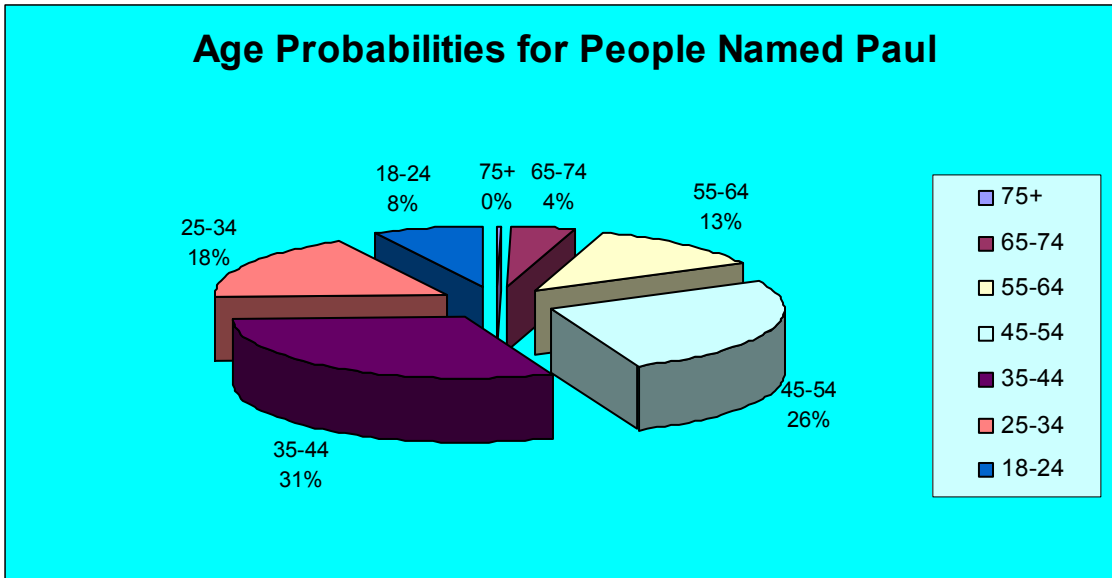
The database contains the following information:-

- a) First Name
- b) Gender (M/F)
- c) The probability of the person being aged 18-24
- d) The probability of the person being aged 25-34
- e) The probability of the person being aged 35-44
- f) The probability of the person being aged 45-54
- g) The probability of the person being aged 45-54
- h) The probability of the person being aged 65-74
- i) The probability of the person being aged 75+

For example, the name Paul has the following information associated with it:-

- Gender is Male (M)

and



This is saying that an estimated 31% of all adults called Paul are aged 35-44, around 26% aged 45-54 and so on. The figures show that the name is most common amongst the middle aged and that few elderly or younger people are named Paul.

3. Applications

a) Profiling

Simply match the name file to the profile source by name and overlay the seven probabilities onto each matched record on the profile source. To estimate the proportion of the profile source within each of the seven age bands; the probabilities within each age group are simply summed across all records with values and divided by the number of records to generate the estimates. This is valid because, for any given name the seven age group probabilities always sum to one. The age estimates for the profile source can then be compared to known national figures (see below) and indexed in the usual way.

b) Targeting

Identify names most associated with a particular age group for list rental purposes. Rank names by the size of the probability within the age band in question and identify those names containing the highest proportions. The probabilities for different bands can be added together before ranking. For example, if you wished to identify people aged under 35 then the probability of someone being aged 18-24 can be added to the probability that they are aged 25-34 and the combined probability can then be used to rank names.

c) Statistical Analysis

The probabilities can be treated as explanatory variables in any statistical exercise such as response analysis or propensity modelling. Techniques such as multiple regression, logistic regression and CHAID can all be used. One of the most recent applications of name information was to try and predict the age of all electors on the Electoral Roll using the name probabilities (as the first name is available) and what was known about them from the Electoral Roll itself (such as Length of Residence, Household Composition, Number of Adults at the address, Presence of Attainers etc.) This exercise proved to be very successful and the name probabilities were essential in accurately predicting an electors age.

5. Expected Match Rates

In compiling the name to age predictors some names did not occur enough times to allow probabilities to be estimated with any degree of certainty. These names do not appear in the file. The file contains information for 3232 names. Something like 97% of the population in Great Britain have a name for which age probabilities have been estimated.

6. National Figures

2007 Population Estimates give the following age and gender distributions for all of those aged 18 or more in the UK:-

Age	Female %	Male %
75+	11.75	7.95
65-74	10.80	10.36
55-64	14.97	15.32
45-54	16.37	16.98
35-44	18.92	19.73
25-34	15.87	16.94
18-24	11.32	12.72
Pop in (000's)	23982.3	22554.8

To calculate indices in any profiling process the above could be used as the base. In modelling and analysis the above could be treated as probabilities and given to those records where the name did not match and no name probabilities were matched to the record, this would allow all records to be used in the analysis process.

David Griffiths

The Research & Analysis Consultancy

1.1.2009