



What is Origins?

September 2018

Origins is a segmentation system which classifies consumers according to the part of the world from which their forebears are most likely to have originated.

Each consumer on a customer file can be placed into one of 250 different 'Origins' types based on their personal and family names. The segmentation could be used for example to identify people on a customer file whose ancestry is most likely to be from Ireland, Italy, Albania or Myanmar.

Origins uses the same information to code customers based on their most likely language and religion. An age estimate and gender can also be appended to the records of most customers based on their names.

A : What information is used to build Origins?

The Origins classification is built from a global file containing the personal and family name combinations of some 601,000,000 adults from around the world. In addition, we have access to personal and family name frequencies covering another 666,000,000 adults. These billion adults are resident in 26 different countries.

Using this information, we have been able to establish the likely Origins code for some 4,000,000 different family names and some 850,000 personal names.

The source of these names is shown below:

	Country	Count	Combinations	Family names	Personal names
1					
2	Australia	11,304,082	x	x	x
3	Brazil	146,303,353	x	x	x
4	Denmark	4,274,408	x	x	x
5	Finland	378,173	x	x	x
6	Germany	64,664,703	x	x	x
7	Greece	184,690	x	x	x
8	Ireland	3,077,097	x	x	x
9	Italy	30,359,327	x	x	x
10	Netherlands	8,257,541	x	x	x
11	Norway	2,889,297	x	x	x
12	Romania	4,032,320	x	x	x
13	South Africa	60,230,874	x	x	x
14	Spain	11,820,566	x	x	x
15	Sweden	4,300,059	x	x	x
16	United Kingdom	47,695,794	x	x	x
17	USA	201,362,132	x	x	x
18	Belgium	9,089,379		x	x
19	France	56,434,686		x	x
20	Austria	2,392,322		x	
21	Canada	3,500,853		x	
22	Czech Republic	3,906,229		x	
23	Hungary	2,166,724		x	
24	India	434,286,400		x	
25	Japan	47,179,928		x	
26	New Zealand	599,381		x	
27	Poland	3,328,966		x	
28	Total	1,164,019,284	601,134,416	1,164,019,284	666,658,481
29					
30					

B : How does Origins work?

In order to code individual customers, Origins makes use of a table which contains information on over 850,000 personal names and over 4,000,000 family names. Each of these names has been examined in such a way as to identify the Origins type to which it is most likely to belong. This evaluation makes use of a number of criteria including the Origins codes of the surnames held by bearers of each personal name, and vice versa; the geographical concentration of the name both within and between countries; the Mosaic codes in which the names are mostly found; and the appearance of diagnostic letter sequences.

This evaluation also establishes the confidence with which we can say a particular name belongs to a particular Origins type.

Looking at the codes associated with both the personal name and the family name and taking into account the confidence level of each, Origins identifies the Origins type to which each customer name is most likely to belong.