BIG DATA AND RESEARCH DATA

THE BEGINNING OF A BEAUTIFUL FRIENDSHIP

Joe Catling • Rupesh Patel

ESOMAR
ESOMAR Office:
Atlas Arena, Azië Gebouw
Hoogoorddreef 5
1101 BA Amsterdam
Phone: +31 20 664 21 41
Fax: +31 20 664 29 22

Email: customerservice@esomar.org
Website: www.esomar.org

Publication Date: November 2016
ESOMAR Publication Series Volume S375 Big Data World 2016
ISBN 92-831-0291-6

COPYRIGHT

Published by ESOMAR, Amsterdam,
The Netherlands
Edited by: Deborah S. Fellows

ABOUT ESOMAR

ESOMAR is the essential organisation for encouraging, advancing and elevating market research worldwide.

With more than 4,900 members from over 130 countries, ESOMAR's aim is to promote the value of market and opinion research in illuminating real issues and bringing about effective decision-making.

To facilitate this ongoing dialogue, ESOMAR creates and manages a comprehensive programme of industry specific and thematic events, publications and communications, as well as actively advocating self-regulation and the worldwide code of practice.

ESOMAR was founded in 1948.

ABOUT ESOMAR MEMBERSHIP

ESOMAR is open to everyone, all over the world, who believes that high quality research improves the way businesses make decisions. Our members are active in a wide range of industries and come from a variety of professional backgrounds, including research, marketing, advertising and media.

Membership benefits include the right to be listed in the ESOMAR Directories of Research Organisations and to use the ESOMAR Membership mark, plus access to a range of publications (either free of charge or with discount) and registration to all standard events, including the Annual Congress, at preferential Members' rates.

Members have the opportunity to attend and speak at conferences or take part in workshops. At all events the emphasis is on exchanging ideas, learning about latest developments and best practice and networking with other professionals in marketing, advertising and research. CONGRESS is our flagship event, attracting over 1,000 people, with a full programme of original papers and keynote speakers, plus a highly successful trade exhibition. Full details on latest membership are available online at www.esomar.org.

CONTACT US

ESOMAR
ESOMAR Office:
Atlas Arena, Azië Gebouw
Hoogoorddreef 5
1101 BA Amsterdam
The Netherlands
Tel.: +31 20 589 7800

Email: customerservice@esomar.org
Website: www.esomar.org

BIG DATA AND RESEARCH DATA

THE BEGINNING OF A BEAUTIFUL FRIENDSHIP

Joe Catling • Rupesh Patel

INTRODUCTION

The world's supply of data is increasing rapidly, and companies hold more data about their customers than ever before. However, this data is by its nature limited in scope: typically to demographics, behaviour such as spending and consumption patterns, and potentially modelled data based on (for example) postal address.

Research can provide a wealth of additional psychographic information around context, emotions, drivers, response to new concepts, and so on – but we can only ever hope to speak directly to a fraction of customers or prospects.

One of the most crucial questions facing both researchers and their clients is: what is the best way to integrate data from the two sources? How can we marry 'big data' and primary research?

Any approach to this challenge will inevitably involve compromise, and when it comes to segmentation, there will always be a trade-off between creating the 'perfect' solution and being able to target segments in practice.

However, where sufficient customer data is available, 'reverse' segmentation represents an alternative to more commonly used clustering techniques, and comes with several distinct advantages.

THE CHALLENGE

With a mature customer base and a propensity model that had outlived its usefulness, Sky's reward-based 'Introduce A Friend' referral scheme had achieved relatively low recent uptake. Furthermore, referrers typically fit a particular profile, as opposed to being broadly representative of Sky's entire customer base.

This led Sky to believe that there may be customer groups for whom a different referral approach might be more successful – perhaps offering an alternative reward, or even playing in a different emotional 'territory' altogether. Identifying these customers, and the propositions that would attract them, could allow Sky to open new 'headroom' and thereby increase referrals.

Once candidate propositions had been established, Sky wished to create a referral scheme segmentation that fulfilled three key criteria:

1. Appendable to Sky's database with 100% reliability, i.e. to be defined solely on internal database fields. This would allow the team to begin targeting the segments almost immediately, rather than (at best) requiring a mapping exercise that would inevitably be time-consuming and somewhat inaccurate, or (at worst) having nice-sounding segments that would be impossible to reliably target.

2. Meaningful and practical, enabling a clear and distinct contact strategy for each segment. (This should, of course, be true of any segmentation.)

3. Straightforward enough for the broader team to understand without the need for additional developmental work. Our solution would represent a compromise between pure statistical rigour and commercial practicality: it was essential that the segments were intuitive for all stakeholders.

METHODOLOGY: SCOPING, IMMERSION, AND IDEA GENERATION

We began with a scoping exercise, involving desk research, interviews and a semiotic review, to understand referral schemes across a wide range of sectors. Following this, we could then categorise referral schemes into three broad groups:

- 'Reward: The clear and simple benefit of getting something back in exchange for doing something – 'you scratch my back and I'll scratch yours'. (Most referral schemes we examined fell into this category, including the current Sky offer.)
- 'Discovery': The passing on of insider knowledge or a well-kept secret, such as an exclusive offer.
- 'Goodwill': Enabling the sense of generosity and satisfaction you get from doing a good deed – for instance helping someone out, or donating to charity.

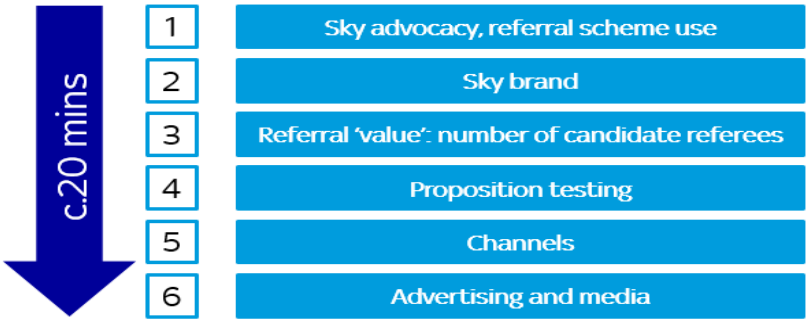
We then ran a qualitative workshop involving focus groups interspersed with discussions among Sky stakeholders to generate, shortlist and refine potential referral scheme propositions. This enabled us to take seven propositions forward into the next phase, including the current offer.

FIGURE 1. PROPOSITIONS TESTED

METHODOLOGY: THE SURVEY

We surveyed 2,462 Sky customers who had not used the referral scheme via an online survey.

FIGURE 2. SURVEY FLOW



From the survey, we identified the 'key questions': the handful of questions on which we wanted our segments to show variety. These were:

- Often use refer / introduce a friend offers
- Often recommend products and services to others
- Aware of current Sky 'Introduce A Friend' offer
- Would recommend Sky to others
- Number of candidate referees: close people, without Sky, who might be interested
- Proposition: appeal
- Proposition: claimed use

METHODOLOGY: DATABASE VARIABLES

We then appended a wide range of database variables to our survey data. From these, we identified 19 variables that could potentially be used to define the segmentation. These included:

- Demographics: age, region, affluence, income, lifestage, household composition, residence type
- Behaviour: tenure, spend, products, bundles, premiums
- Existing models: other segmentations, the current propensity model

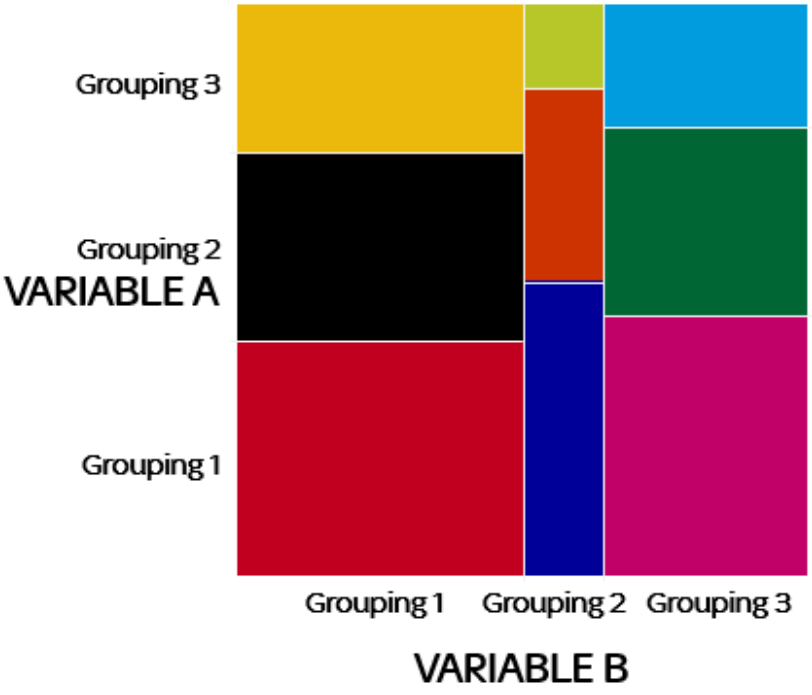
It should be noted that, while some of these database variables were modelled rather than entirely accurate, for our purposes this didn't matter – we would evaluate the suitability of these variables as they stood, rather than using them as proxies for survey responses. By contrast, we could instead have asked about (for example) household composition in the survey, and used responses here to define our segments. We would then have required a modelled variable for household composition within the database, and any inaccuracy in this modelled variable would have made the segment allocation itself less accurate.

Given that we wanted all segments to be a reasonable size, we combined values in each of the 19 database variables into two to three 'groupings' per variable, with each grouping ideally containing at least 20% of the sample. This was done primarily using common sense (for example, banding adjacent age groups), but where appropriate was also informed by responses to the key questions – combining database values with similar responses to keep groupings as internally homogenous as possible.

METHODOLOGY: EVALUATING THE "LENSES"

We then paired each of our 19 variables, as well as considering each variable individually, to create 190 potential 'lenses', each of which contained up to 9 (i.e. 3 X 3 groupings) 'target units'. For example, within the lens 'age X income', one target unit was 'aged 46+ and income £40k+'.

FIGURE 3. EXAMPLE TARGET UNITS



Our next challenge was to identify which of the 190 lenses had target units whose responses to the key questions saw the most variety. In particular, it was crucial that target units had different proposition preferences. As expected, the current Sky proposition was the most popular overall, but we could see that many potential target units preferred alternatives. (At this stage, we were able to eliminate three of the seven propositions: not only were they less popular overall, there were almost no target units who preferred them. This left us with four: Current Offer 1, Rewards 3, Goodwill 6 and Goodwill 7.)

For each lens, for each of the key questions, we calculated the standard deviation as if all respondents within each target unit had given the mean response for that target unit. The greater this figure, the greater the variety the lens showed across the key questions.

We then weighted and summed these standard deviations to establish a single metric for overall 'variance' across the key questions.

FIGURE 4. KEY QUESTION STANDARD DEVIATION WEIGHTINGS

Variable description	Often **use** refer / introduce a friend offers	Often **recommend** products and services to others	**Aware** of current Sky offer	Would **recommend** Sky to others	**Candidates:** close people, without Sky, who might be interested	Viable propositions: **appeal**	Viable propositions: **use**
Weight assigned	10%	10%	10%	10%	10%	25%	25%

This allowed us to rank the 190 lenses, and thereby establish a shortlist to take into the final analysis phase – though as it happened, the segmentation we ultimately chose was generated by the lens with the highest variance of all.

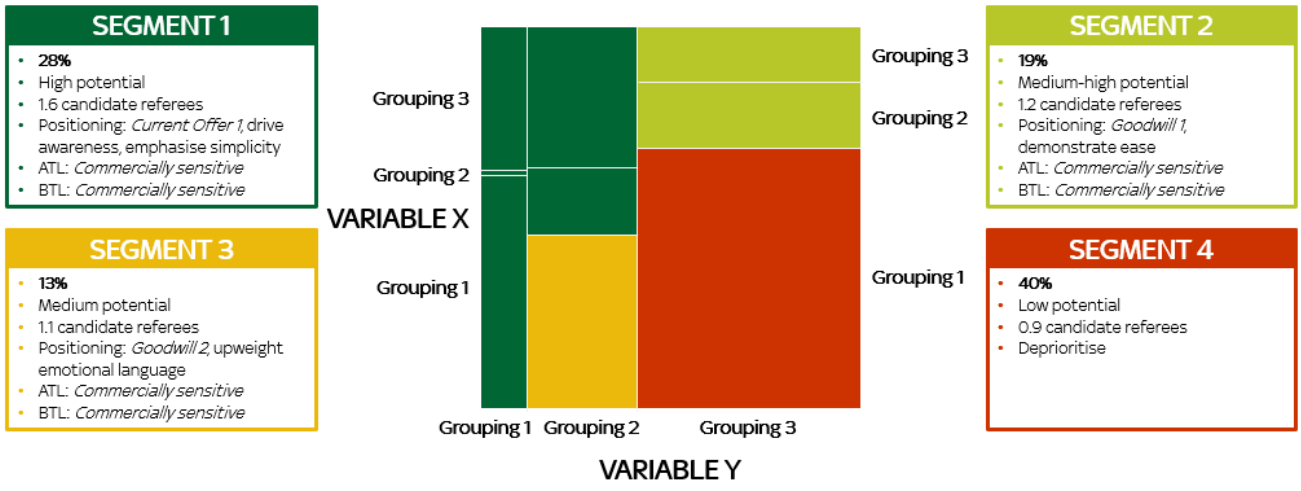
METHODOLOGY: RECLUSTERING

Each of our shortlisted lenses contained up to nine target units. However, it was generally felt that nine segments would be too many to be of practical use, and in some cases target units were very small. We therefore 'reclustered' target units, i.e. joined them together. This was driven primarily by the need for coherence: we joined 'adjacent' target units according to responses to the key questions – trying to keep segments as internally homogenous as possible.

As stated previously, it was essential that the segments were practical and intuitive, as well as displaying different proposition preferences, and it was on this basis that our candidate solutions were evaluated. Our chosen segmentation contained three segments who each preferred a different proposition, along with a segment we recommended ignoring altogether, thus enabling more efficient deployment of resources.

Other survey questions added profiling information, along with informing specific contact strategies for each segment. And although for commercial confidentiality reasons we haven't shared them here, we gave the segments suitable names.

FIGURE 5. SOLUTION



EARLY RESULTS

At time of publication, we have only the results of an initial marketing campaign. However, tailored comms have shown an uplift in response of 28% for segment 2 and 24% for segment 3 (i.e. the segments for whom a different approach was required), demonstrating the effectiveness of the study. Further refinements will hopefully result in even greater improvement.

RECOMMENDATIONS

Clearly, reverse segmentation is only possible where a client has a customer (or prospect) database. It may also not prove appropriate when said database:

- Contains only a limited subset of customers, especially if they are not themselves representative of the broader customer base, e.g. loyalty card holders.
- Contains relatively little information: in the Sky example detailed above, for instance, had we only had access to simple demographic data, our solution would not have been as powerful.
- Is too small for a robust base to be obtainable, given expected incidence rates.
- Has considerable gaps.

In such cases, clustering based only or largely on survey responses may be the only viable approach.
It should also be borne in mind that, given the tools that currently exist, reverse segmentation is likely to be considerably more time-consuming than other clustering techniques. Furthermore, when it comes to attitudinal profile or concept response, the segments may not be as distinct as those generated by other means.

However, reverse segmentation has the distinct advantage of enabling reliable appending to the client database. Especially when it comes to below-the-line comms, it circumvents the classic criticism of so many segmentations: 'This looks great, but how do I actually know who these people are?'

Though the precise methodology described here was developed for this study, reverse segmentation in general is not an entirely new approach. However, it lacks the prevalence of other more widely used methodologies, for example factor and cluster analysis (by whatever technique). In a world where clients hold ever more data about their customers, we believe a 'reverse' segmentation ought to be more and more applicable, representing as it does an effective way of combining this data with primary research.

REFERENCES

Jones, Frazier, Murphy, Wurst – Reverse Segmentation: An Alternative Approach – Sawtooth Conference, 2006

THE AUTHORS

Joe Catling is Head of Analytics, Relish Research, United Kingdom.

Rupesh Patel is Audience and Portfolio Controller, Sky UK, United Kingdom.