Identifying audience targeting opportunities utilising behavioural data and third party segment membership.

S. Wallace - Mindshare Worldwide
S. Hamilton - Mindshare Worldwide
E. Danika - Mindshare Worldwide
K. Saba - Mindshare Worldwide
V. Cook - Mindshare Worldwide

21 September 2018

Abstract

Traditional segmentations rely on proxy and claimed information when identifying media targeting opportunities. Mindshare has partnered with a 3rd party data supplier to augment survey data with behavioural and targetable 3rd party segment information so that the targeting information is based upon actual behaviour as opposed to claimed behaviour. This data augmentation allows for the creation of segmentations with direct buyable components to be created. With information around 3rd party segment membership the translation into a buyable strategy becomes more streamlined as well as providing a level of information prior to a campaign that can have an impact upon optimisation strategies due to a higher level of knowledge surrounding websites visited and potential audience overlap.

Introduction

Whilst there is some disagreement and exaggeration around how many adverts consumers are exposed to each day a conservative estimate would place the number somewhere between 500 and 1000. (http://blog.telesian.com/how-many-advertisements-do-we-see-each-day/)(http://www.bandt.com.au/marketing/consumers-exposed-600-messages-day-getting-search-right-rules-marketers)

In this sea of noise it has never been more important for a brand to deliver the right message to the right person in the right time and place. However this is not an easy task, according the Neilsen-Norman Group the length of time most users spend on a website is 59 seconds; given the amount of adverts a user sees daily if they aren't engaged with the content on a website placing your message there is a misallocation of funds. This is why being able to identify the way in which a target audience behaves is so crucial to an effective targeting strategy and why behavioural targeting is somewhat of a holy grail in media planning. When constructing a media targeting strategy it begins with one simple question: Who are the people I want to reach with my message?

In order to answer this initial question an audience segmentation is performed, typically the data upon which we construct our audiences is obtained from a bespoke survey in order to identify audiences relevant to the campaign and client objectives. The questions asked and the data we can obtain from such a method allow us to construct a demographic, geographic, or psychographic segmentation as well as a combination of these. Whilst effective in answering the initial question "Who are the people I want to reach with my message?" unless there are questions regarding media activity in the survey we may not have a way in which we can reach them. Even with these questions survey data is not a perfect source due to issues with respect to misrepresentation, sampling, and bias. Whilst answering online is less prone to misrepresentation, there is still an issue with regards to claimed vs. actual behaviour as has been well documented in previous research. (http://www.uxbooth.com/articles/bridging-the-gap-between-actual-and-reported-behavior/) This is why obtaining data around actual user behaviour would be immensely valuable when constructing a targeting strategy as we could identify from our sample areas where we know our desired audience will be.

Such data is difficult to obtain, however there are companies that have access to this information beyond the big names within the industry. These companies (data providers) analyse the behaviour of a user and group them into third party segments which are available for purchase in buying platforms. These third party segments are then bought by traders when activating a media targeting strategy, however problems exist in this phase too. If we describe our audience in terms of demographic profiling whilst these can be bought they are often wide reaching, and what is often used to narrow down a demographic profile is information such as psychographic data. Whilst this narrows down targeting, there are not always direct analogues within buying platforms and proxies are made by traders in order to best realise the targeting strategy. Another problem can be that the information used to construct the audience segmentation, and as such the eventual description of the audience, has no targetable components when buying.

This highlights the underlying problem faced when activating a media targeting strategy, even if we have identified the right person and the right message if we cannot reach them in the right place our strategy is not as effective as it could be.

In response data providers have constructed an offering of "Data enrichment" where they connect with a data management platform (DMP) in order to map the data within and onto their information so it can be bought. However, this process is typically opaque and dependent upon a DMP; moreover it is a step at the end of the audience profiling and can require certain types of data for this to be perfomed. Typically this process is also only available if you have first party activity data and the data provider is able to match that user in their database, meaning that if you dont have this information available or a large enough data sample this cannot be done effectively. Whilst there are ways in which we can append third party information to survey data it still does not provide us with behavioural data or a direct buyable link.

Having identified this problem Mindshare have partnered with a 3rd party data provider in order to produce a solution that as far as we are aware is the first of its kind. This paper outlines the way in which we have been able to append behavioural data directly to survey responses enabling us to construct an audience segmentation using survey data that has direct buyable components built using behavioural data.

Methodology

The aim of this research was to show the way in which Mindshare and a 3rd party data supplier can partner on a survey in order to construct a targeting strategy for our created audiences based upon the information obtained from the data enrichment.

A 2000 respondent survey was run in France around the mindsets a respondent holds when making purchasing decisions with demographic and psychographic questions. Upon consent a 3rd party data supplier placed a tracking pixel in order to monitor their behaviour across websites for a period of 120 days, returning to us the websites visited and the 3rd party data supplier segment they would be placed into the segments to be purchased are given as numbers for anonymisation purposes. This information from 1292 respondents was then appended to the survey responses for further analysis.

We constructed an audience segmentation based upon responses to a survey based around purchasing habits for the purposes of this research in order to identify directly targetable opportunities based upon the 3rd party data augemntation. In order to construct the audience segmentations we utilised unsupervised machine learning techniques utilising an algorithm Mindshare have developed which investigates k-means, k-mediod, and NMF across a range of potential clusterings and chooses the audience segmentation with the most distinctions across the psychographic questions selected and a set of questions in which we want to see distinctions. The size of the audiences are:

Segment	1	2	3	4	5
Size	39.7%	18.6%	14.5%	16.4%	10.8%

Data privacy and GDPR

Mindshare commissioned the survey to be done in partnership with a 3rd party data supplier utilising a panel provider GMI Lightspeed. Prior to the survey the consent text was shown to the potential respondent outlining the fact the cookie would expire 120 days after it was placed and where they could visit for more informaiton

The pixel placed was used solely for collection of data and although the respondent was modelled into the 3rd party data segment they were not placed within, meaning the respondent was exposed no additional adverts as a result of their participation in the survey.

Results

As a result of the segmentation utilising the data provided by a 3rd party data supplier we are able to construct an activation strategy for psychographic audiences that is built upon observed behaviour and memberships obtained from behavioural activity. This direct targetability of psychographic audiences directly addresses complaints surrounding such segmentations and provides us additional information about our respondent's behaviours. Utilising this approach means that we can build a strategy that can be readily activated using the AMP which is Mindshare's programmable targeting strategy that enables us to realise a client's strategy without dilution constructing rulesets traditionally infeasible to manage. This level of granularity requires a high level of data quality that this approach provides.

Website activity

Using our 5 segment psychographic audience segmentation built upon behavioural data we can look at the websites the user has visited using the enriched data from a 3rd party data supplier. First let us look only at the top level domain and filter out any domain that accounts for less than 5% of the audience traffic. We can see in the figure below a word cloud where we scale the top level domain by the proportion of the audience that has visited the domain and colour the text by how many of our audiences have visited this top level domain.

We can see from Figure 1 that the domains that have the highest percentage of users are also the ones which are shared between audiences, whereas websites with smaller reach are more likely to be unique to an audience. This information can be used to split whitelist targeting where there is a higher likelihood of seeing the desired user, which is useful when you have different creative messages for different audiences.

Let us look now at the average number of visits each user within an audience makes to websites.

From Figure 2 we can see that on the most visited websites such as "priceminister.fr" and "toner.fr" have different average views per audience. Let us quickly look to see whether we can use information such as time of visit and total number of visits to a site to identify patterns audience membership based upon the activity from "www.priceminister.com"

From Figure 3 we are able to identify which audience is most active during a time of day with audience 1 being the most active during the hours of midnight to 3am, then audience 2 being the most active from 4am to 7am, with other audiences being active during the core hours. Some of this activity will be due in part to the size of the audiences as Figure 3 is the percentage of activity per hour, so each hour will sum to 100%. Let us use the χ^2 residuals to identify times when each audience is the most active in comparison to its own activity as well as the activity of the other audiences.

Looking at Figure 4 we see that the smaller audiences 3, 4, and 5 are relatively more active in the afternoon. This information can be activated in a targeting strategy by placing more importance on impressions served in the afternoon on "www.priceminster.com" with these audiences as relative to the other audiences and itself this is when the audience is more active.

Figure 1: Top Level Domains

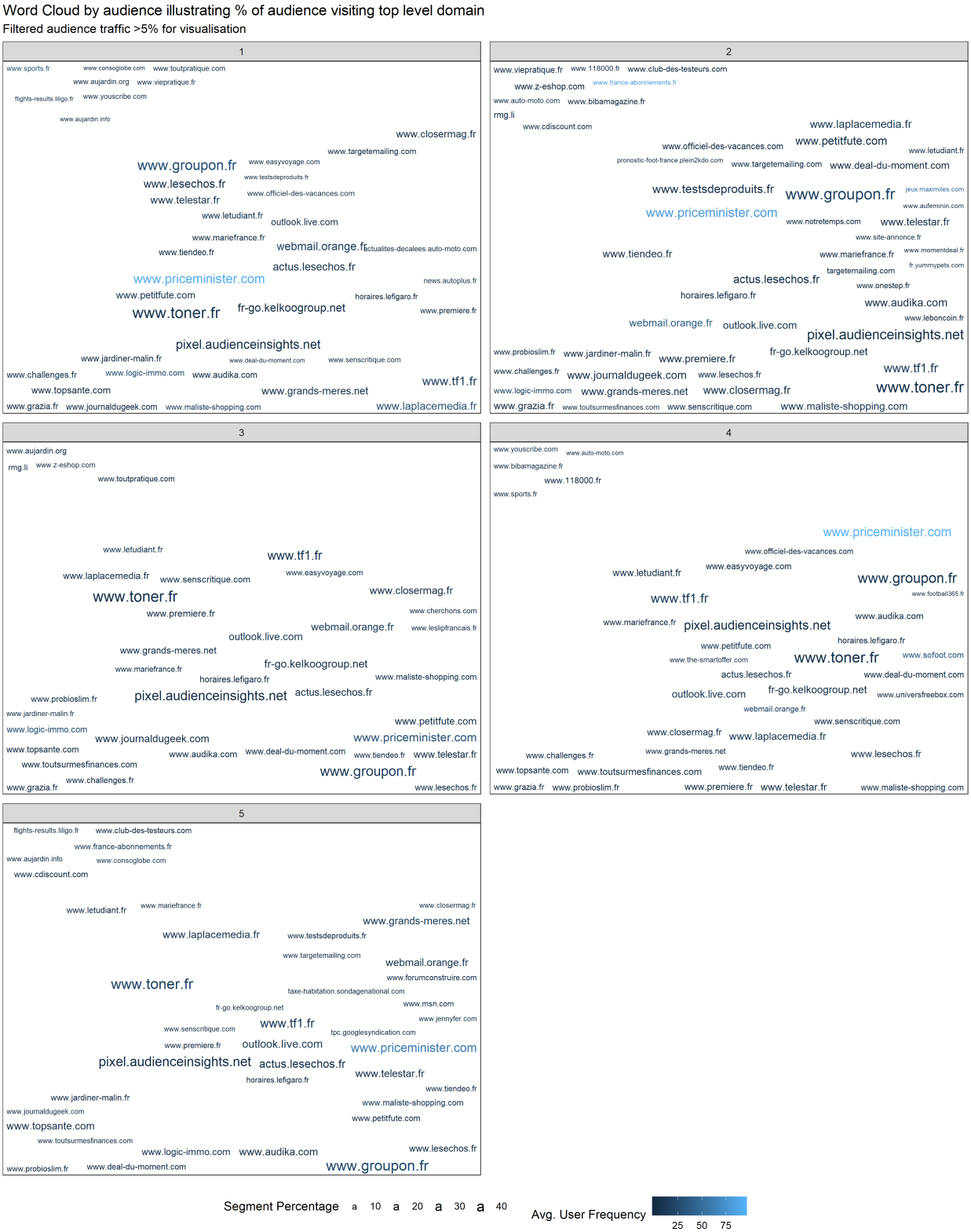

Figure 2: Top Level Domains - Average visits

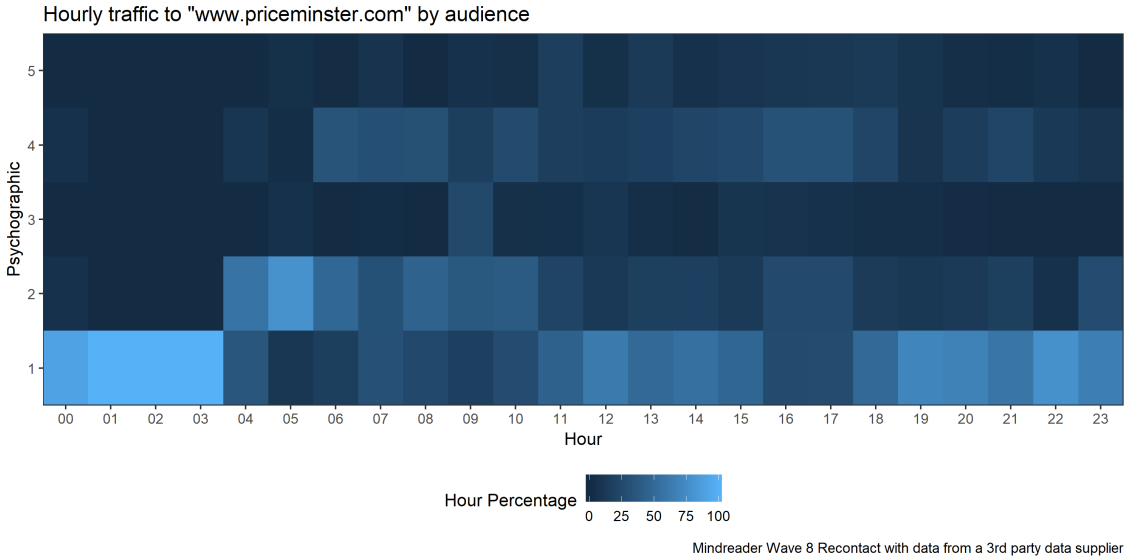

Figure 3: Percentage breakdown of traffic per hour

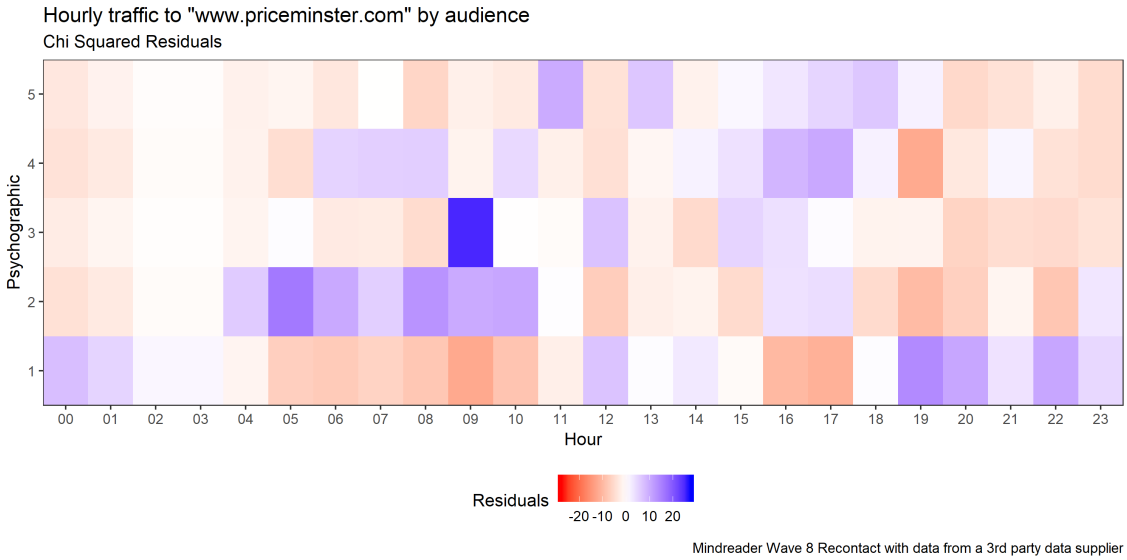

Figure 4: Breakdown of traffic per hour per audience using χ^2 residuals

Whilst we don't necessarily know how many times during a live campaign a user has visited an external website this shows that we could use information that we do have, such as the number of adverts served, to identify likely audience membership based upon time of activity. This information can inform strategic planning, creative messaging, and can be incorporated into a high detail activation strategy that can be realised using the AMP. With this information we can amend our activation strategy to take into account the likelihood of the active user belonging to one of our audiences, changing messaging or bidding according to the importance of each audience.

With the information regarding websites that user has visited we can construct a whitelist on which we should place a higher bid, a general whitelist which we can learn more about as the campaign progresses, and a granular strategy based upon a audience's behaviour. These insights provide us the ability to add an additional granular level to traditional activation strategies.

3rd Party Data Segments

Not only can we analyse website activity using this approach we can identify the 3rd party segments that our psychographic audiences belong to and as such should form the core of our targeting strategy from the information provided by a 3rd party data supplier. Let us now explore how we can identify 3rd party segment membership for our psychographic audiences.

Similarly to how we assessed websites in the previous section we will now look at a word cloud of 3rd party segment membership, filtering out any 3rd party segment where less than 1% of the audience is a member for visualisation where the size is the percentage of the audience that belongs to this segment, and the colour is the number of audiences that have people in this segment.

What we see from Figure 5 is that 3rd party segments that are unique to an audience are only reaching around 1%-2% of the audience. This homoegeneity in 3rd party segment membership presents an interesting challenge to buying audiences using 3rd party data segments, when we look at the respondents in our survey each of them belongs to an average of 13 segments, given that our audiences have not been built around 3rd party data segment membership having this homogeneity is to be expected as it would in any response to other questions in survey data. The way in which we can identify targeting opportunities will be to assess differences from the baseline, we will do this using χ^2 residuals and the resulting plot can be seen in Figure 6. We have arranged the 3rd party data segments in Figure 6 in order according the dominant audience with respect to the χ^2 residuals to illustrate the targeting opportunities where if we were to purchase this 3rd party data segment we are more likely to reach our audience.

Let us look at an audience in detail to see how we can understand how purchasing these 3rd party segments could aid us in reaching our desired audience. We will look at our second Psychographic audience and only the 3rd party data segments where they are the dominant audience, Figure 7 shows us this information and that where they are the dominant segment the other audiences aren't as present according to the χ^2 residuals. This approach is similar to double indexing in survey data, and allows us to identify targeting opportunities taking into account the homogeneous nature of 3rd party segment membership.

From Figure 7 we can see that for this audience we have around ten 3rd party segments before the residuals for other segments become high indicating a larger than expected membership, this shows us that these few 3rd party segments might give us a better chance at reaching our audience without overly diluting our targeting with other audiences. Let us look at the unique reach we might expect from adding these 3rd party segments in the order on the graph - Figure 8 - and the breakdown of our audiences that would be purchased at each step in which we have marked when the total breakdown of our desired audience becomes less than 50% - Figure 9.

We see from Figures 8 and 9 that as we increase the number of 3rd party data segments we purchase the reach of our desired audience improves at the cost of dilution as we begin reaching more of the other segments at the same time. When the total number of anticipated people within our audience we theoretically purchase is still above 50% we can buy 9 segments with an expected audience reach of 14.5%. This tradeoff is anologous to lookalike modelling when building an audience from information contained within a DMP and could

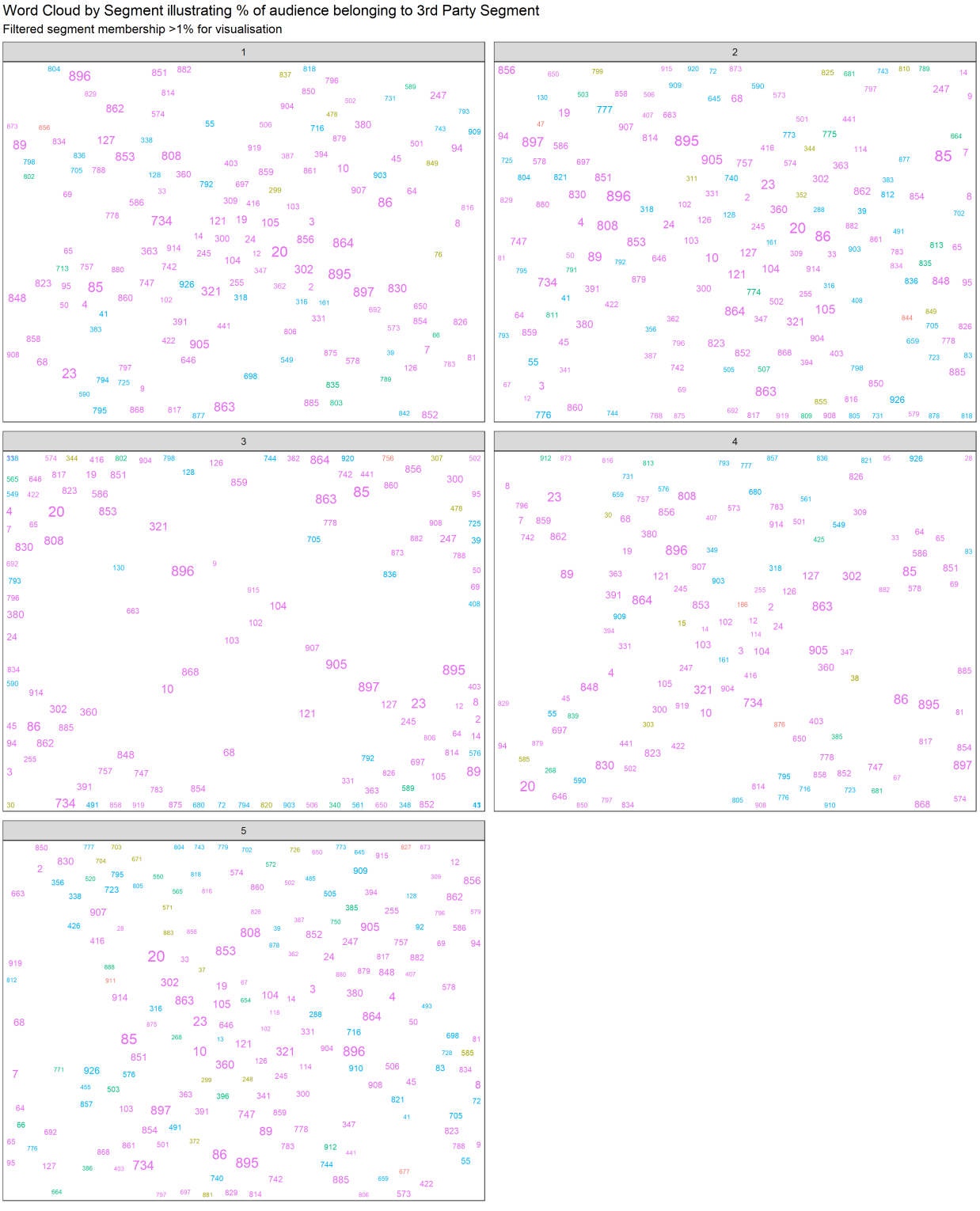

Figure 5: Segment membership by audience

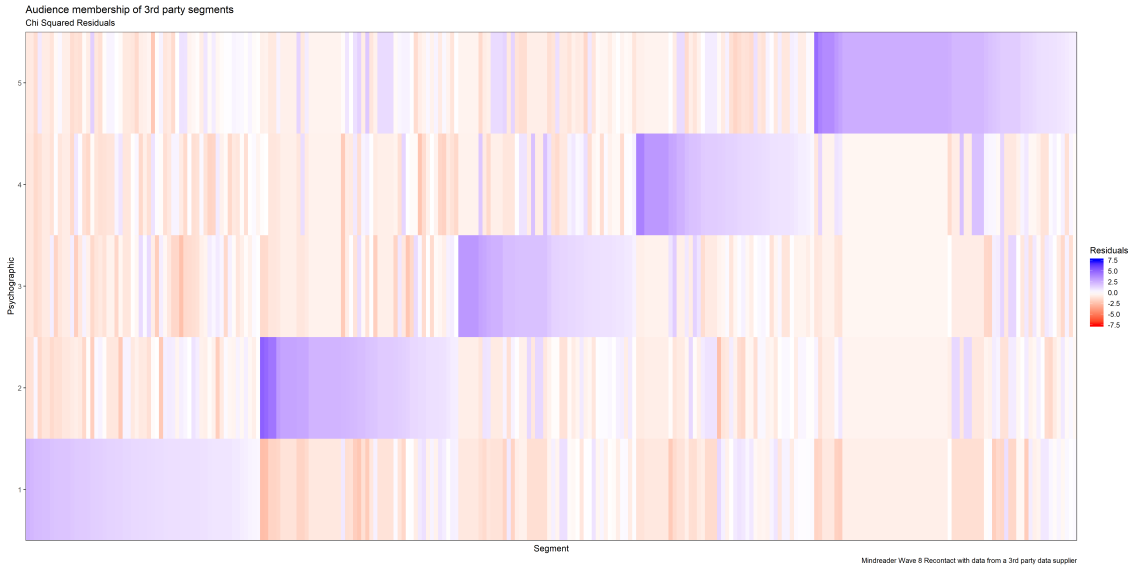

Figure 6: Breakdown of segment membership per audience using χ^2 residuals

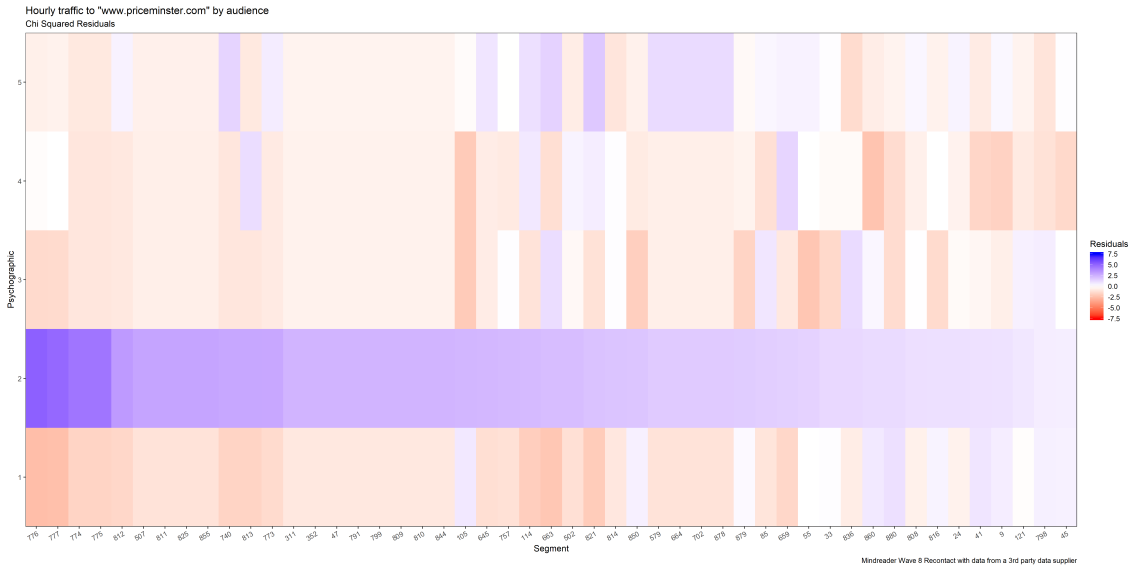

Figure 7: Breakdown of segment membership for psychographic audience 2 using χ^2 residuals

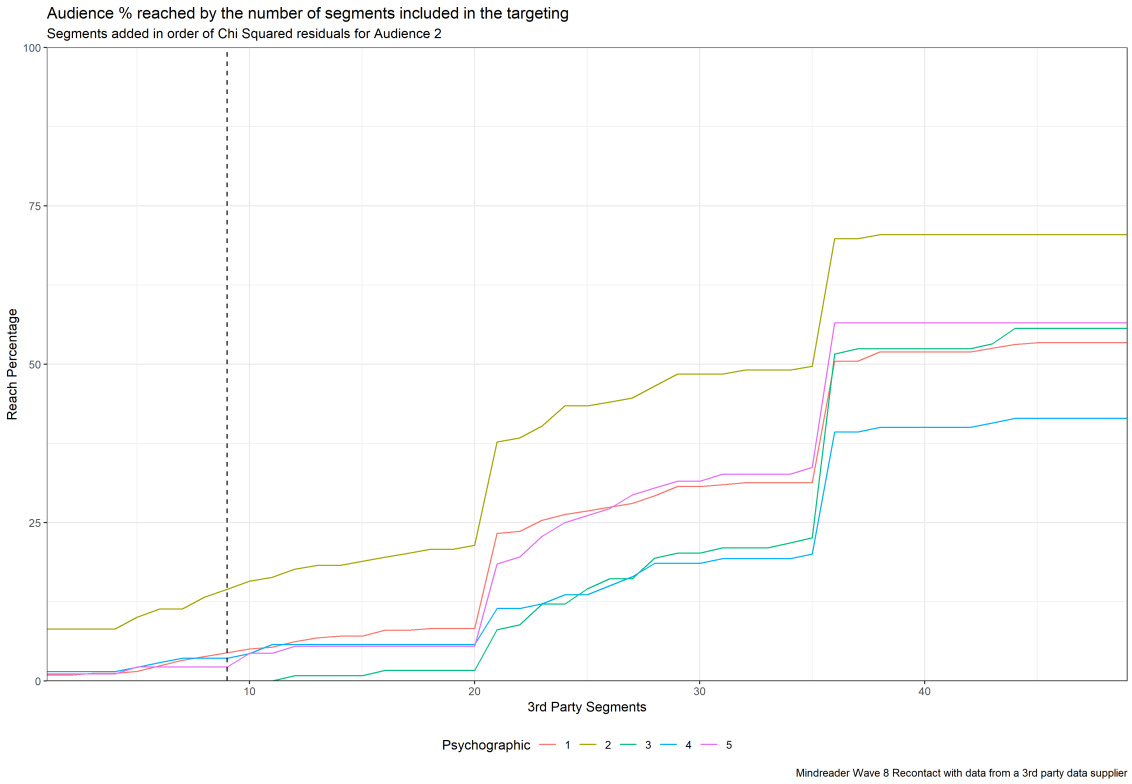

Figure 8: Percentage of audience reached with number of 3rd party data segments included

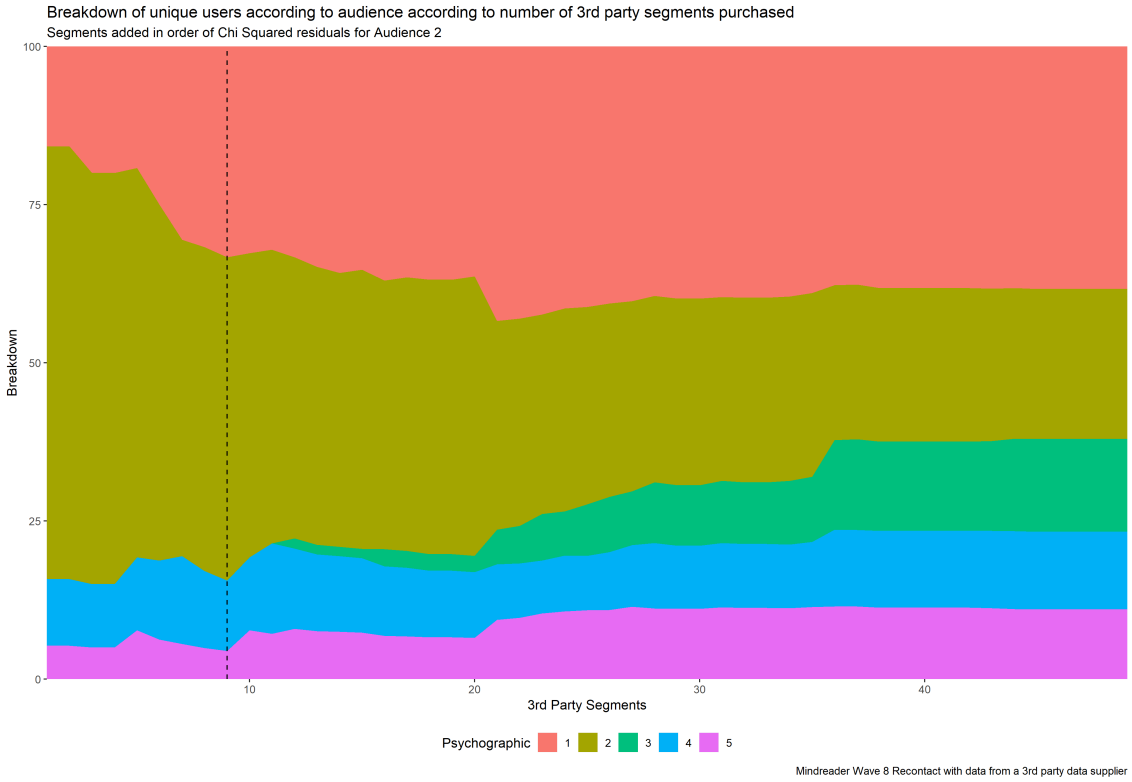

Figure 9: Breakdown of audiences reached with number of 3rd party data segments included

be used in a similar manner understanding that as we increase the audience reach using 3rd party data segments we will decrease the accuracy of our purchasing including users outside of our desired audience. Within Figure 9 we can see that as we include more 3rd party data segments the breakdown of audiences purchased becomes more similar to the overall segment composition of the audiences within the sample.

What we find from this research is that there is a limit to purchasing 3rd party data segments in order to reach our audience, and that a better way to reach our desired audience may well be to look at behavioural information and when they are more active. However, augmenting such a strategy with 3rd party segments could be an additional strategy to identify areas of possible improvement with respect to our campaign KPIs.

Whilst we can still construct a prospecting ruleset we are able to identify the strongest direct links between our audience segmentation and the 3rd party data segments available in DSPs. This approach not only allows us to build a campaign ready to be activated using the AMP but we can also utilise the data to inform a traditional targeting strategy.

Discussion

This research allows us to understand how we can activate audiences in a more effective manner thereby improving ROIs and KPIs for our activation strategies.

An industry wide problem faced when constructing a targeting strategy containing 3rd party data segments is potential overlap, and therfore optimising towards 3rd party segment membership. With the AMP we can construct rulesets to take overlap in 3rd party segment membership into account however it increases exponentially with the number of 3rd party data segments we utilise. Moreover with targeting 3rd party data segments we cannot affect the overlap of audience membership within these purchasable segments. With this approach we can gain an idea of any potential overlap before the first impression is served, grouping together such segments or creating a ruleset that accounts for the overlap as appropriate. What we have also shown is that when purchasing 3rd party data segments we should limit our purchases as when we include too many we start to mirror the overall composition thereby potentially negating the benefit of audience targeting. With this approach we are able to identify the ideal tradeoff point where we can get the benefit of these 3rd party data segments without fully diluting the reason for purchasing such segments. This then allows traders to focus upon optimising towards segments that provide the best return on investment for our key performance indicators, balancing the tradeoff between scale and audience accuracy when purchasing these segments. Additionally part of the problem with audience construction using survey data is the chance of bias in responses due to false information being provided by the respondent. Through constructing 3rd party data segment lists in this manner we can identify such responses and if these responses are nor used in the segmentation adjust our definitions accordingly.

Future Research

This research also raises an interesting discussion around utilising 3rd party data segments for prospecting given the homogeneity that can be seen across segments, and as such several approaches could be taken. The first is to conduct a test using 3rd party data segments to target individual audiences or whether they should only be used to narrow the potential users and media behaviours used to subset audiences within activation strategies. Second would be to see how effective conducting an audience segmentation would be using the 3rd party data segment membership and whether that is suitable for planning and strategic purposes the same way a traditional audience segmentation is, we could also try a similar approach but using only website activity.

One future avenue would be to try and utilise the information prov.ided by 3rd party data suppliers to obtain results that are similar to lookalike modelling based upon information contained within DMPs. In order for us to do this we would look to append information from a range of 3rd party data suppliers therby allowing us to create potential cookie pools across a range of suppliers thereby widening our potential reach. This has

a potential benefit of increasing the scale we can get using 3rd party data targeting whilst understanding the segments which best suit our goals and objectives. By utilising more than one data supplier it creates the opportunity to understand our activation strategies in greater detail as well as improving the way in which we activate audiences. With this approach we can also have greater control over how the "lookalikes" are obtained as we can look directly at who is potentially being targeted through the information appended in the survey data.

Finally given enough data points we could look at using machine learning to try and build a predictive model to further refine our activation strategies based upon behavioural information.

Conclusion

In this research we have shown how Mindshare can partner with a 3rd party data supplier to augment survey data with behavioural information and 3rd party segment membership. We have shown how we can use this information to augment and improve upon traditional activation strategies through a better understanding of where we can activate our audience segmentation. We can identify times of day on websites where we would expect to see each audience and which 3rd party segments are more likely to contain our audience. This information can be incorporated into traditional programmatic activation strategies and in higher granularity activation strategies using the Mindshare's AMP. The value of this approach is that before the first impression is served in a campaign we have a greater level of information around the third party data segments that we are purchasing as well as websites that our sample audience has visited. This information allows us to create a targeting strategy that can be readily activated using the AMP or traditional means and has the potential to increase campiagn effectiveness due to a higher level of information in campaign creation.