Beyond Demographics – Targeting Likely Consumers through Psychographic Traits

Authors:
Steven Millman, Chief Scientist, Simmons Research
Pat Pellegrini, President & Chief Research Officer, Simmons Research
Hu Yang, Senior Lead Statistician, Simmons Research
Chris Nay, Lead Statistician, Simmons Research
Tamara Barber, Senior Manager, Simmons Research

Abstract:

As advertisers and market researchers, we live in a world of demographic targeting, trained to identify our likely consumers and audiences from broad traits such as age, gender, language, geography, and the like. Data on such traits are widely available, but tend to be overbroad and necessarily coarse. For example, the demographic profile for comic book readers and paintball enthusiasts are probably very similar. An advertiser looking to reach either group might prepare a target of white, lower income males, age 18-34. For advertisers of steel-tipped boots, however, this demographic profile is not nearly granular enough. Avid comic book readers are likely to have very different interests in steel-tipped boots than the more typically outdoorsy paintball players. Psychographics on the other hand, which assess customers' opinions, attitudes, and interests, tell a far richer story. Targeting against simple additional psychographic traits such as "I like to be outdoors" would allow the marketer to clearly distinguish between those groups and achieve a far greater return on investment (ROI) with their ad spend by more precisely focusing on their prospective buyers. The data mining techniques that define Simmons' Predictive Consumer Insights bring to bear on the nearly 1,000 psychographic variables in the National Consumer Survey provide the ability to create much more precise and efficient targets, identify the psychographic profile of brand/media consumers, and to track changes in that profile over time.

Predictive Consumer Insights – the Methodology:

The Simmons National Consumer Survey (NCS) is a rich, nationally representative, probabilistic paper survey produced by Simmons Research that brings in over 25,000 respondents annually. Included in the NCS are thousands of questions related to the use of specific brands and types of media consumption. Also in the NCS are nearly 600 psychographic attributes and a wide array of demographic information. To these variables, we also add RMT's 265 DriverTags™, which are themselves a form of psychographic identifier derived from media viewing patterns. DriverTags are words that were distilled from over 13,000 words comprising all of the psychological words in the English language which were, in turn, derived from the inspection of all the words in the English language. These 265 words were found to have the highest correlation with a subscriber becoming a loyal viewer of a never-watched-before program that had been recommended by the artificial intelligence of a platform. Taken together, they reflect the hidden, underlying psychological motivators of program choice. Each respondent in the NCS is assigned DriverTags based on the tags associated with the shows they self-report having consumed.

Predictive Consumer Insights is a statistically robust data mining approach that identifies traits that can be shown to be predictive of either brand choice or media consumption, as well as determining the independent effect of each trait. This process is based on a series of nested regression models, in which each iteration further reduces the set of potential predictors until they have been distilled to a set of statistically significant psychographic variables. Because brand use or media consumption are dichotomous variables, taking on only the values of zero (non-user) and one (user), a logistic regression approach is employed. Logistic regression predicts the probability that the value of a dichotomous variable will be one, meaning that the action occurred, whether that be a brand purchase or viewership of a media channel. The following equation defines the logistic regression in which y is the predicted probability of the event occurring, x is a predictive variable, and β is the regression coefficient which indicates the degree to which the predictive variable is moving the probability of the event. This can include any number of predictive variables and in this general equation is identified as up to j variables.

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j)}}$$

Because there are nearly a thousand potentially predictive variables between the demographics, psychographics, and DriverTags, it is not possible to simply run a regression with all of the variables at the same time. To solve this problem, the psychographic variables and DriverTags were broken up into smaller sets, organized around common themes or identified using factor analysis. Factor analysis is a statistical technique that identifies sets of variables whose responses tend to be related. In every model, the demographics are also included as control variables. The number of variables in each model is different, but for the sake of understanding the methodology, let’s assume there were exactly 50 variables per model and 20 models for a total of 1,000 psychographic models.

When the first model is run, there will be a number of collinear variables identified. Collinear variables are variables that are so highly correlated that they are essentially the same with respect to predicting the likelihood that a respondent will use a brand or consume a kind of media. These are easily identified by high standard errors and correlation. Once removed, the model is rerun, and again examined for collinearity. Once collinear variables are removed, psychographic variables not significant at the 95% confidence intervals are removed. This distilled set of reduced variables is then reserved, and the next model of 50 variables is run. This process is repeated for each of the 20 models. Consider the simple four-model case below, where each color is a different variable model with related potentially predictive psychographics.

$$\begin{tikzcd}
y = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_{50}x_{50})}} \arrow[d] & y = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_{50}x_{50})}} \arrow[d] & y = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_{50}x_{50})}} \arrow[d] & y = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_{50}x_{50})}} \arrow[d] \\
x_2, x_{12}, x_{21}, x_{28}, x_{39} \arrow[d] & x_7, x_9, x_{15}, x_{41}, x_{42} \arrow[dl] & x_1, x_7, x_{14}, x_{33}, x_{49} \arrow[dll] & x_{11}, x_{18}, x_{26}, x_{30}, x_{35} \arrow[dlll] \\
y = \frac{1}{1+e^{-(\beta_0+\beta_2 x_2+\beta_{12}x_{12}+\beta_{21}x_{21}+\beta_{28}x_{28}+\beta_{39}x_{39}+\beta_7 x_7+\beta_9 x_9+\beta_{15}x_{15}+\beta_{41}x_{41}+\beta_{42}x_{42}+\beta_1 x_1+\beta_7 x_7+\beta_{14}x_{14}+\beta_{33}x_{33}+\beta_{49}x_{49}+\beta_{11}x_{11}+\beta_{18}x_{18}+\beta_{26}x_{26}+\beta_{30}x_{30}+\beta_{35}x_{35})}} & & &
\end{tikzcd}$$

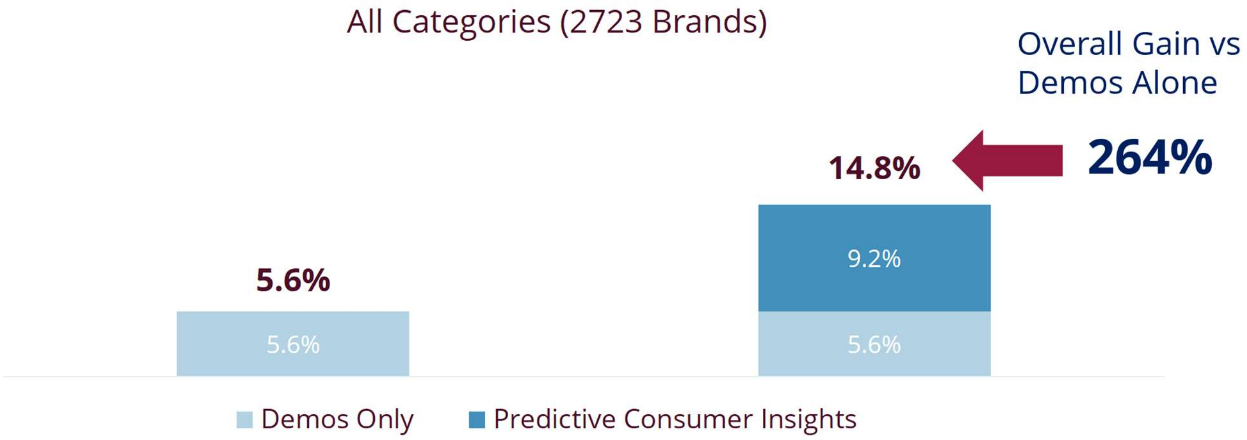
In this last step, a model is run using all of the retained explanatory variables, once again along with the demographic variables (not shown in above models). The processes of removing collinear and non-significant variables is then repeated, until a final set of predictive variables is achieved.

$$y = \frac{1}{1+e^{-(\beta_0+\beta_{21}x_{21}+\beta_{28}x_{28}+\beta_{39}x_{39}+\beta_9 x_9+\beta_{15}x_{15}+\beta_{42}x_{42}+\beta_1 x_1+\beta_{11}x_{11}+\beta_{18}x_{18}+\beta_{30}x_{30}+\beta_{35}x_{35})}}$$

This final predictive model can then be used to provide a likelihood score for every NCS respondent that is related to the probability that an individual will be a brand user or media consumer. Each brand takes on average approximately 150 regression models to produce a final predictive model. This process has been automated and repeated to date for 2,723 brands and television shows, which were selected based on a conservative minimum incidence rate of 1,000 consumers from among the 25,000 in a typical NCS full-year file. In total, almost a half million regression analyses were performed.

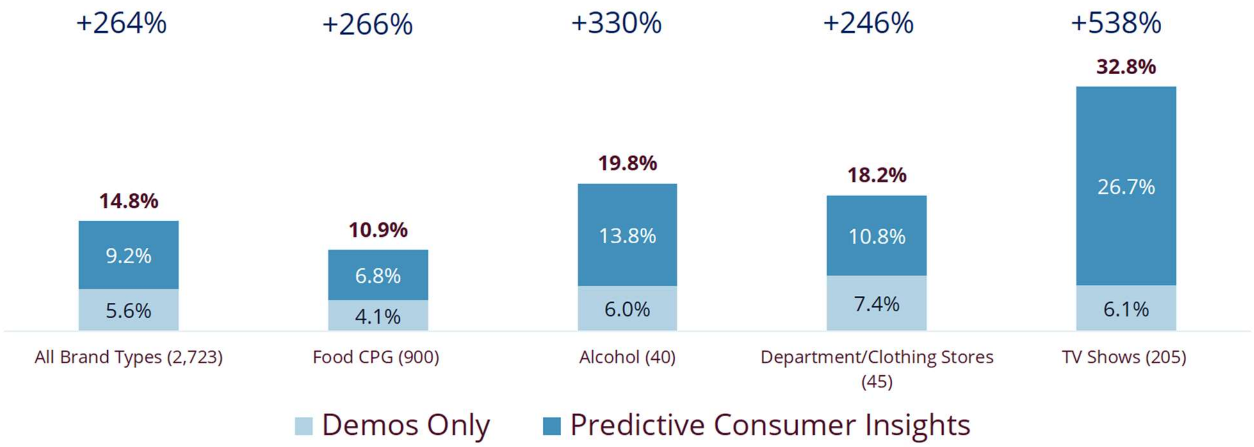
To validate the benefits of this approach, we wanted to see if these models provided more explanatory power than using demographic variables alone. One of the metrics that can be produced in a logistic regression is called a pseudo-R^2. The pseudo-R^2 is similar to an adjusted R^2 in a linear regression, and refers to the percent of variance explained by the model. The value of an R^2 varies from zero, or no explanatory power, to one, meaning that the model is completely deterministic. While the pseudo-R^2 is not a perfect measure of explanatory power, it does provide a reasonable way to compare the effectiveness of various approaches. Figure One describes the effectiveness of using the psychographics in these models as compared to demographic variables alone.

Figure One: Explanatory Value of Psychographic Variables vs Demographic Variables Alone



Across all brands, demographic variables provide very limited explanatory power, on average explaining only 5.6% of the variance in brand choice. Adding psychographic variables adds greatly to this explanatory power, increasing the average percent of variance explained to 14.8% – an increase of 264% over demos alone! Figure Two below breaks this down further by brand category.

Figure Two: Explanatory Value of Psychographic Variables by Industry Category

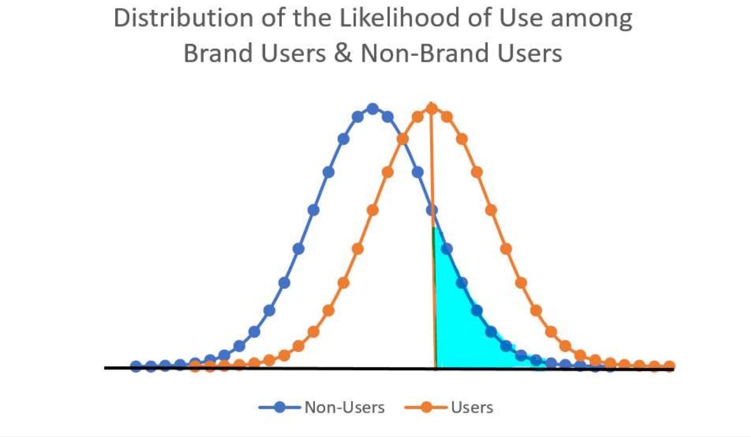


The value of psychographic variables in predicting brand use is very consistent across all brand categories measured by the NCS. Figure Two shows a sampling of those brand categories. In each case, there is at least a three-fold increase in explanatory power when psychographic variables are used rather than demographics alone. With the exception of TV shows, DriverTags provide a small but significant increase in explanatory power across all brand categories to the nearly 600 psychographic variables used in the model. Because DriverTags are related to the psychographic traits driving program choice, it is unsurprising that they provide much of the additional explanatory power that this methodology adds to predicting television show viewership. On average, Predictive Consumer insights predicts 32.8% of the variance in TV show viewership, more than five times the power of demos alone.

Use Case One: Targeting the Most Likely Prospects

Once a final predictive model has been generated that predicts the likelihood for any individual to be a consumer of a brand or a viewer of a TV show, it becomes possible to compare the distribution of those likelihoods among consumers/users and non-users. The natural assumption is that the average likelihood of a brand users is higher than the average likelihood of a non-user, an assumption that has been uniformly been reflected in the data. Figure Three shows an example of the distribution of likelihood of brand use among brand users and non-brand users.

Figure Three: Overlap of the Likelihood of Brand Users and Non-Brand Users



It is highly valuable for an advertiser to be able to target non-users of their brand in order to increase reach or market share. That said, the most valuable non-users are those who would be most likely to use the brand, a task that Predictive Consumer Insights is uniquely able to accomplish. In Figure Three, the distribution of the likelihood of NCS user and non-user respondents are overlapped. Having looked at thousands of brands, we have found that a useful definition of high-value prospects are those whose probability of brand use are at least as high as the top half of the actual brand users. This region of high-value prospects is highlighted in light blue. This group of brand prospects becomes a segment in the NCS, and can be targeted efficiently across any traditional media channel (television, print, radio, etc.), or can be loaded into a variety of activation platforms for precisely targeted digital ad campaigns. In addition to this, the competitors to your brand can also be identified, locating those individuals most likely to be persuadable to switch to your brand.

Use Case Two: Fine Tuning the Message

Additional critical outputs of a logistic model are the coefficients that tell you the independent effect of each predictive variable, holding constant the effect of all of the other variables. Why is this so important? Imagine you are in a city or county and you've commissioned a study to understand how to reduce costs from fire damage, exploring the factors that predict the amount of damage fires have been doing. If you simply look at variables one at a time, for example through indices, you might see that fires that result in significant property damage index very high on the intensity of the fire, the value of the property on fire, and the number of firemen that respond to the fire – and all at approximately the same index. This obviously does not mean all three are causal factors of fire damage, nor does this mean that that they have the same impact even though they have similar index values.

In a regression, you are be able to examine each of the following as they relate to predicting fire damage in dollars:

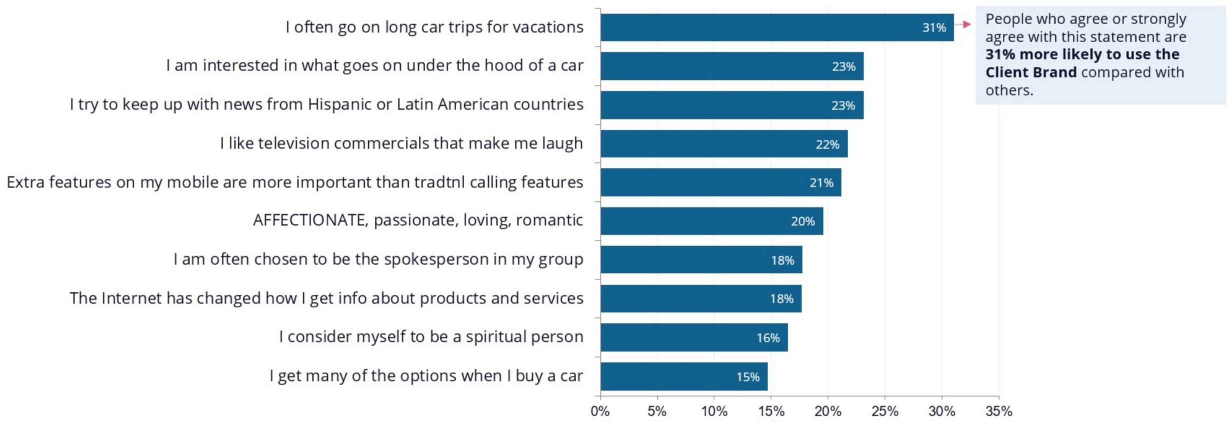
a) The impact of fire intensity, holding constant building value and fire response
b) The impact of fire response, holding constant building value and fire intensity
c) The impact of building value, holding constant fire intensity and fire response

Doing this you would see that (a) and (c) have a real impact on costs associated with fire damage, whereas (b) does not. This is because the fire response is a function of fire intensity and not the other way around. It overindexes, but has no explanatory power. The municipality cannot really do anything to affect the value of the buildings and clearly reducing the number of firefighters responding will not help to reduce fire damage. As a result, the county or city has learned that it should focus on ways to reduce fire intensity (i.e. fireproofing, strict building codes, etc.) in order to reduce costs and keep fire costs down.

Similarly, a marketer who wants to understand what factors predict potential users for his/her brand might find that – while two different attitudes overindex for brand use – these attitudes might be so highly correlated with one another that only one of the two really matters. Advertisers need to know this so that they aren't spending a lot of money and energy messaging to both attitudes in their targets.

The regression coefficients in a logistic regression are more challenging to interpret than in linear regressions where the coefficient is simply the increase in Y for every increase of one in X. Fortunately, a simple mathematical conversion of logistic coefficients produces a value called an odds ratio. An odds ratio is the percent increase in Y based an increase of one in X. The coefficients generated in the final brand models allow an advertiser to see not only all of the variables that drive the likelihood of using their brand, but also their relative strength. Figure Four describes the ten strongest predictors of an alcohol brand.

Figure Four: The Top Ten Positive Predictors of an Alcohol Brand



People who agree or strongly agree with this statement are **31% more likely to use the Client Brand** compared with others.

The way to interpret the odds ratios here is that a person who agrees with the psychographic statement, "I often go on long car trips for vacations" is 31% more likely to be a user of the client brand than someone who did not. For this brand, they may have already known, or created, the association with road trips or that their brand is popular among Hispanics, but some of the other psychographic drivers may have been a surprise. Knowing that humorous commercials are appealing to users provides valuable information about the kind of advertising campaign they may wish to produce for highest ROI. The finding that likely users identify themselves as spiritual and affectionate can also provide direction to creating the most powerful advertising campaign.

This analysis, as mentioned before, can also be repeated for each of a brand's competitors. Overlapping predictors define the category, more than differentiating the brand, and so will be of lesser value to the advertiser. Non-overlapping predictors provide both a clear understanding of the psychographic makeup of a brand's existing user base, but also a window into the makeup of their competitors. Using these data, a clever advertiser could easily construct a campaign to target category users that currently purchase from their competitors.

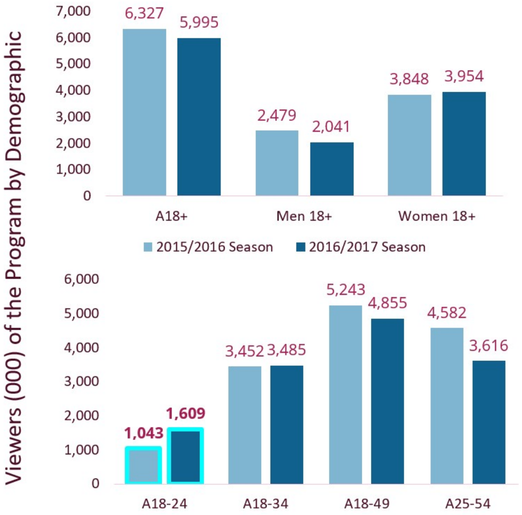
Use Case Three: Changes in the Psychographic Makeup of Your Brand

Another important advantage of these kinds of psychographic analyses is that an advertiser can take the psychographic profile and see what happens to their brand users or prospects over time, much like brands already do with their more traditional brand tracker surveys. Brands or media channels need to understand the changing landscape of their users and prospects so that they can adapt to them, much the same way as they do today with shifting demographics. Occasionally, brands will want to intentionally shift the way they are viewed via their advertising. This can be the result of significant events, such as a brand recall, a public relations issue such as a spokesperson making offensive comments, or simply a reaction to reducing market share. A recent example of this is the Carl's Jr ad campaign. For several years, the Carl's Jr ad campaign had been centered around very salacious imagery, including such spokespersons as Paris Hilton and Kate Upton in bikinis. In March, 2017 a new ad campaign launched in which a storyline was introduced that Carl Jr's dad had returned to the company to get his son and his company back on track by focusing on more wholesome, family-friendly tropes. If successful, the ad campaign should alter the psychographic profile of the typical brand users, and hopefully making it attractive to a larger group of consumers, thus leading to higher market share. The kind of psychographic data mining described in this paper would provide a far more complete picture of the changing user base for the brand than would shifting demos alone.

Another significant use case would be of great value to the makers of serial TV shows. A great deal is made of increasing or declining viewership as these affect the price an advertiser can be asked to spend on ad during the show. Viewership changes are today described nearly exclusively in terms of valuable demographic groups, but there is room for much more precise targeting. A network could, for example, identify all of the brands whose psychological profiles line up well with a particular show. If an automaker's most likely prospects strongly overindex in a show's audience, the network could charge a premium for that space, even if the show did not do as well when considering only the standard audience demographics. Further, a network can observe the psychographic profile of their audience changing season over season. This can be used to evaluate the root causes of lost audience and how to recover, or to track their progress in aligning their audience to psychographic profiles more attractive to advertisers.

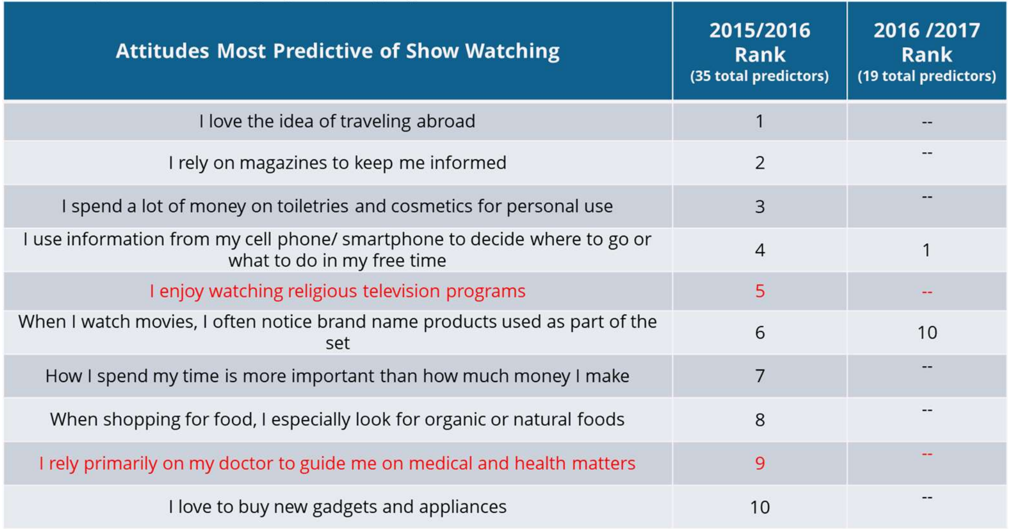
Consider the case of a real prime time network comedy series that lost about 15% of its audience year over year. The network can look at their shifting demographics (Figure Five) and see that they had small losses in older viewers and among men, along with a slight improvement in young adults, insufficient to make up for the losses. These data don't give the advertiser much to work with when thinking about how to improve for the next season and bring back their audience.

Figure Five: Changes in Audience Demographics Year over Year



Changes in the psychographic makeup of the audience is more instructive. Figure Six shows the top ten strongest predictors of viewership for this show in the 2015/2016 season (negative predictors are in red). As you can see, the psychographic profile of this audience has changed far more than can be identified from the small changes in demographics. If the show were, for example, attracting viewers who enjoy foreign travel and has lost them, a future season might incorporate exotic locations in order to help bring them back. In this show, several of the main characters have become less financially unstable and more career focused, which may be leading to the loss of the part of the audience for whom how much money they make is less important than how they spend their time. The network can also see that the audience has maintained heavy use of smart phones, a fact that could be used for a multi-screen campaign. Further, the advertiser can continue to argue for premium ad spend, because the audience appears to have maintained a very valuable ad receptivity trait – that viewers notice product placement in movies.

Figure Six: Changing Psychographic Profiles for a Show in Decline, Year Over Year



Attitudes Most Predictive of Show Watching	2015/2016 Rank (35 total predictors)	2016 /2017 Rank (19 total predictors)
I love the idea of traveling abroad	1	--
I rely on magazines to keep me informed	2	--
I spend a lot of money on toiletries and cosmetics for personal use	3	--
I use information from my cell phone/ smartphone to decide where to go or what to do in my free time	4	1
I enjoy watching religious television programs	5	--
When I watch movies, I often notice brand name products used as part of the set	6	10
How I spend my time is more important than how much money I make	7	--
When shopping for food, I especially look for organic or natural foods	8	--
I rely primarily on my doctor to guide me on medical and health matters	9	--
I love to buy new gadgets and appliances	10	--

Conclusion

Taken together, the addition of psychographic elements to brand/audience profiling creates a wide variety of benefits to an advertiser or advertising platform over the use of demographics alone. The addition of psychographics greatly increases the explanatory power in predicting likely users or viewers, typically three to five times more effective than demos alone. The use of modern statistical and data mining techniques on traditional market research data allows brands to let the data speak for themselves, rather than having to attempt to guess what traits might be predictive, saving advertisers time and making their use of available data far more productive. Advertising platforms can use the psychographic profile of their audience to appeal directly to the brands most likely to find prospects on their media channel and to make an argument for premium ad spend. Finally, brands can track their psychographic profiles over time to both measure the success of their efforts to change their branding as well as to take early action if their profiles start to drift away from their desired user of audience composition.