MEMORANDUM

To: The Media Consortium

Cc: Laura Quinn, Zak Clement-Bremer, Clay Haynes, Erik Brauner

From: Jason Ost

Date: September 1, 2009

Re: Prospect model for The Media Consortium

In June and July 2009, pursuant to a Data Evaluation Agreement between the Foundation for National Progress (The Media Consortium) and Catalist, the consortium provided several sets of membership and subscriber data[1] to Catalist. The goal was for Catalist to build a predictive model that may identify an enriched prospect universe of individuals who look similar to current members and subscribers of Media Consortium organizations.[2]

While The Media Consortium's member organizations ostensibly share a common political viewpoint, the media channels are diverse, from websites and magazines to television and radio stations. To leverage the common political stance of the Media Consortium member organizations, Catalist built a look-alike model predicting each individual's interest in an affiliation with media outlets that have a progressive viewpoint, such as members of The Media Consortium. This model was trained by comparing individuals affiliated with two or more Media Consortium members to a random sample of individuals from the voter file, with the aim of finding characteristics that make those affiliated with Media Consortium organizations distinct from the general population.

Catalist attempted several models to distinguish the channel preference of individuals affiliated with Media Consortium groups—such as print vs. online and text-based vs. audio-visual—with no meaningful results. There may be some benefits derived from segmenting interested individuals according to their primary media channel or some other classification, but doing so would require more knowledge about how this model would be employed by the consortium members.

This memo outlines the model Catalist built for the Media Consortium and provides some initial out-of-sample validation for the model's effectiveness.

Look-Alike model for The Media Consortium

The Media Consortium members offer a relatively unique product: media with a specifically progressive political focus. This tends to narrow the potential audience, but from a microtargeting or modeling perspective, this narrowness also makes it somewhat easier to find patterns of characteristics that make individuals affiliated with consortium members (the target universe) different from those who are not (the reference sample).

Catalist built a look-alike model that compared these two groups and found the relevant characteristics that separated them. To enrich the target universe and exploit the large number of datasets provided by the Media Consortium, we chose to only include individuals who have an affiliation with two or more Media Consortium outlets. This still left nearly 200,000 individuals in our target universe, though it is worth noting that the size of the print publication lists were much larger than other lists, leading to a heavy representation of print subscribers in the target universe.

To limit the effect of responsiveness bias—i.e., the tendency for the model to find generic "responders" rather than individuals who specifically would be interested in Media Consortium members—we limited both the target and reference universes to voters who had cast a ballot in the 2008 general election. Since these voters have shown a baseline level of participation in the political process, the model should be able to focus more on what specifically makes people respond to the Media Consortium messages, rather than on what makes them more likely to be politically active in general. This limitation reduced our target universe to 161,000 individuals.

The resulting model was highly predictive when tested on a hold-out sample of records in the target universe and reference sample. The most significant variables in the model paint a broad portrait of Media Consortium supporters as white progressive voters with high education, voting participation, and above-average magazine and internet consumption. Among the dozens of significant variables in the model, Table 1 provides a selected list of the most predictive variables.

Table 1. Selected Variables in the Media Consortium Look-Alike Model

Variable	Direction	Influence
Presidential Preference Model	+	Strong
Educational Attainment Model	+	Strong
Partisanship Model	+	Strong
Environmental Activist Model	+	Strong
African American	-	Moderate
Progressive Donor Model	+	Moderate
Voted in a primary election	+	Moderate
Hispanic	-	Moderate
Subscribes to magazines	+	Moderate
Single female household	+	Moderate
Healthcare Issue Model	+	Moderate
Heavy internet user	+	Moderate

To validate the model, we scored the individuals affiliated with only one Media Consortium member, as well as a 1% sample of the voter file records in the states of interest. Figure 1 shows how the members of the holdout sample's target universe are distributed among the model scores grouped into equal 20-tiles across all voters in the United States. With 59% of the target universe in the top five percentiles (i.e., top 20-tile) of the model scores, and 69% of the targets in the top decile (i.e., top two 20-tiles) of scores, this model appears to be highly predictive among the highest-scoring records.

Figure 1. Distribution of Individuals Affiliated With Only One Media Consortium Member, Among Look-Alike Model Scores Grouped Into Equal 20-tiles Across the U.S.
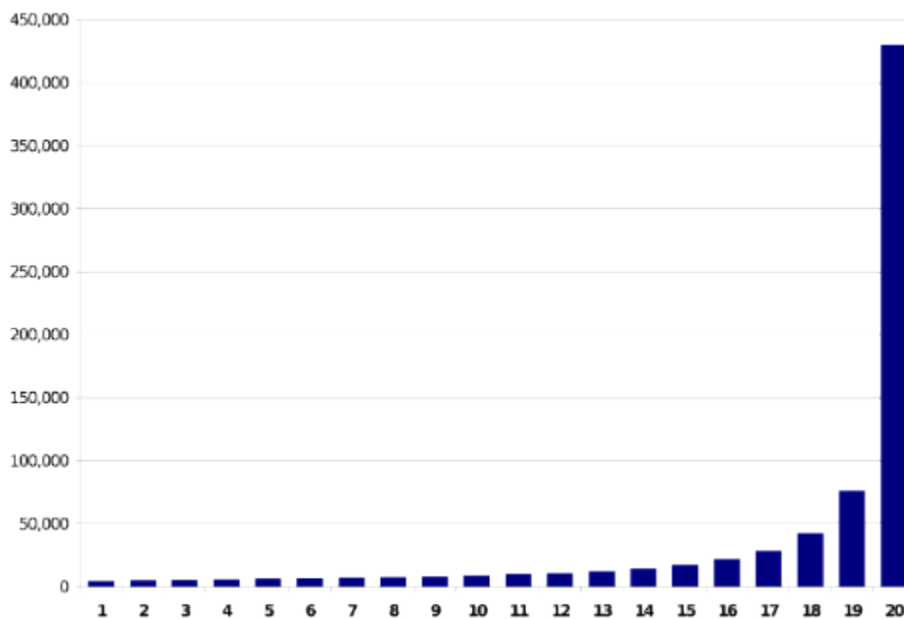

Validation of the Look-Alike Model

Solely for the purpose of further validating the model, Catalist sought "out-of-sample" data sets that may provide insight into whether the model is truly identifying individuals. The most directly relevant source of out-of-sample validation in this instance is a national consumer survey to which Catalist has access that provides information about a range of habits across various media channels.

Although the New York Times is not a Media Consortium member and does not specifically have a progressive political viewpoint, it would not be surprising to find that individuals affiliated with consortium members have a higher propensity to read the Times. In fact, the model shows excellent separation among the 20-tiles on who reads the New York Times' print edition at least once a week, with nearly one in ten individuals in the top 20-tile reading the Times in print at least once a week (see Figure 2).

Figure 2. Share of Each 20-tile of the Media Consortium Look-Alike Model That Reads the New York Times in Print At Least Once a Week
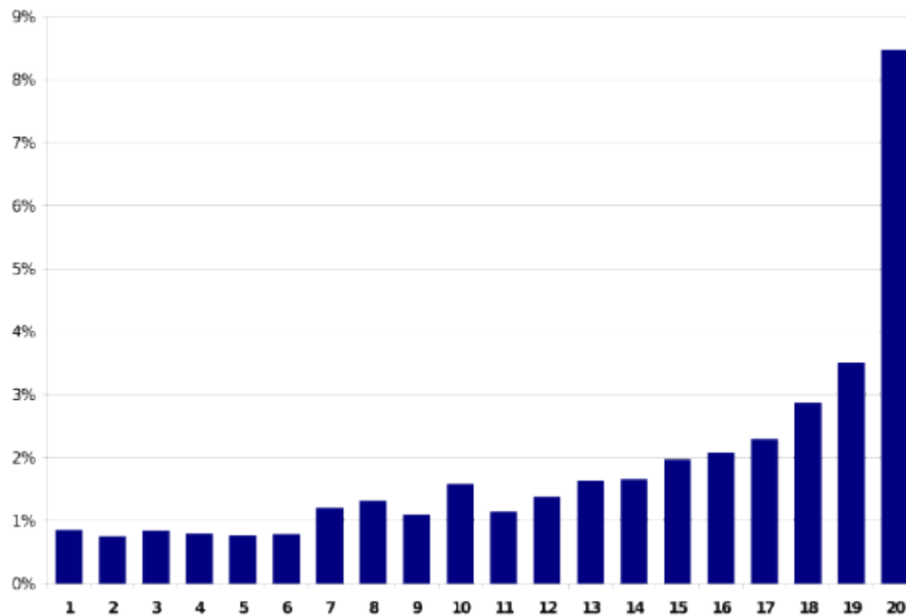

Looking at a media outlet with a specifically non-progressive political viewpoint, FOXNews.com, the model also shows clear separation among those who visited that site in the past month (see Figure 3). This distribution is suggestive that the model is not only predicting those who consume a lot of media, but specifically is targeted towards consumers of progressive media.

Figure 3. Share of Each 20-tile of the Media Consortium Look-Alike Model That Has Visited FOXNews.com In the Past Month
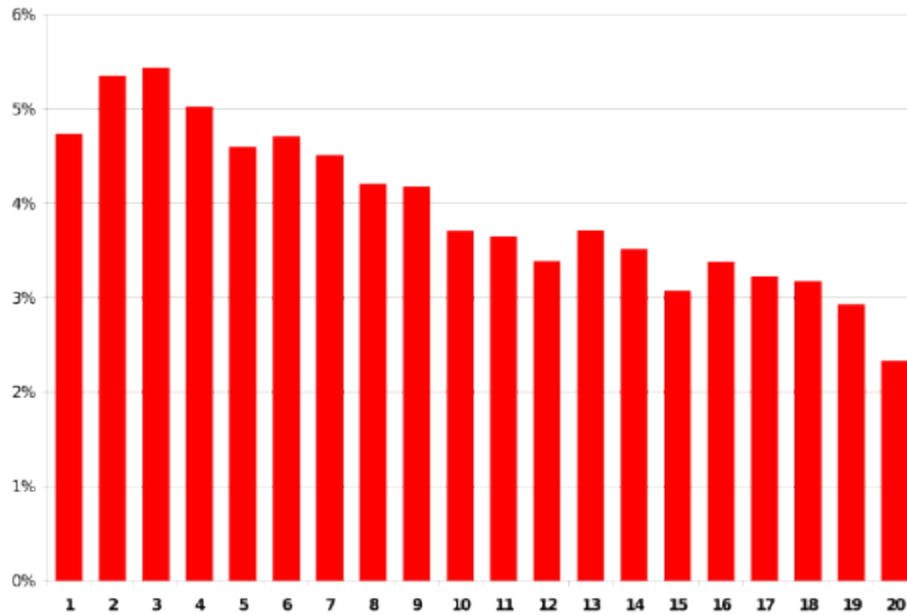
The distribution of USA Today print readers among the look-alike model's 20-tiles is further evidence that the model is targeting *progressive* media consumers, as there is no discernable pattern that indicates a general bias towards reading a widely available, politically neutral publication (see Figure 4).

Figure 4. Share of Each 20-tile of the Media Consortium Look-Alike Model That Reads USA Today in Print At Least Once a Week
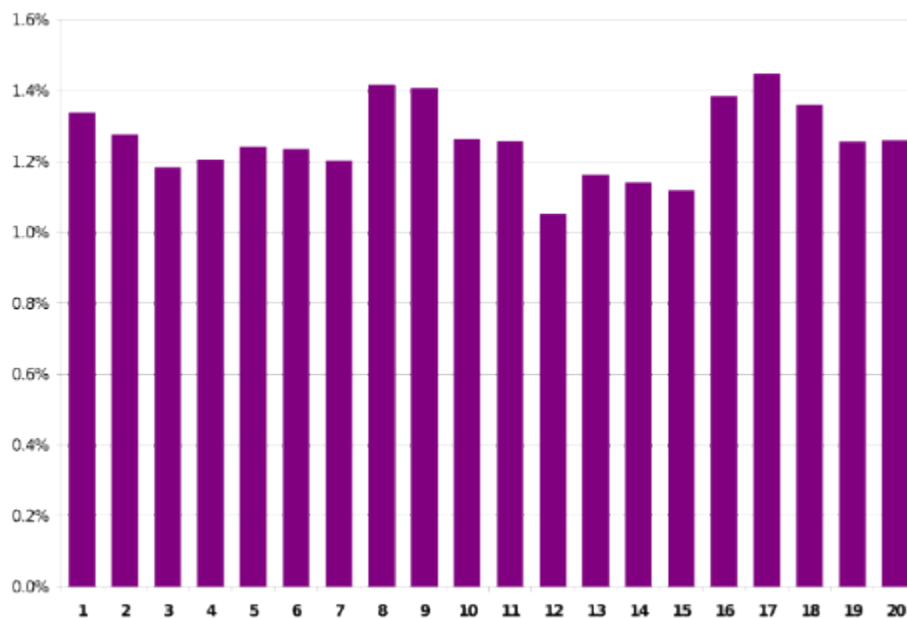
In terms of online media consumption beyond FOXNews.com, the model appears to be somewhat predictive of reading and/or contributing to blogs (see Figure 5). Since the consumer

survey did not separate blogs by content type, it is likely the slight increase at the lower end of the model are conservative blog readers. In general, though, there is not tremendous separation in the model among blog readers, except for the very highest 20-tile.

Figure 5. Share of Each 20-tile of the Media Consortium Look-Alike Model That Has Read or Contributed to a Blog in the Past Month
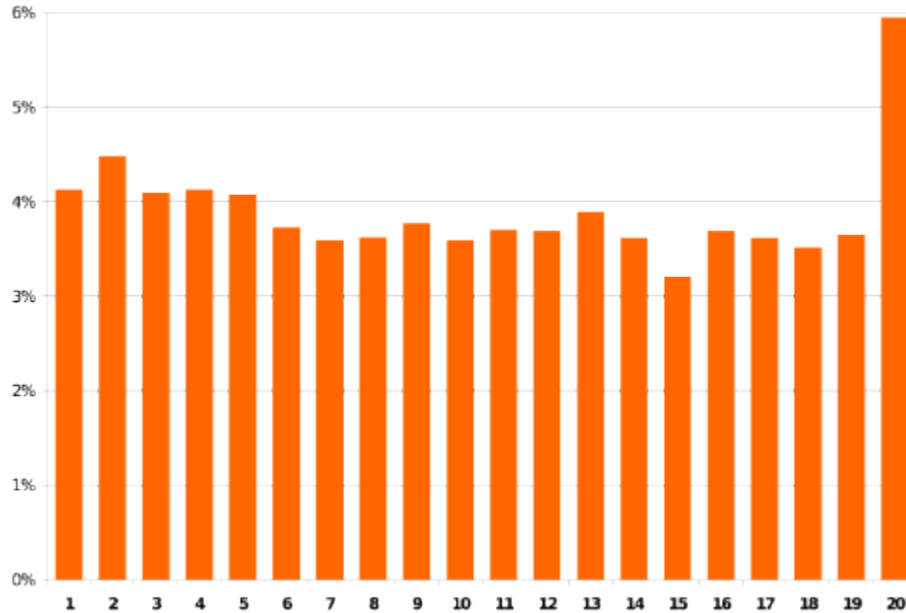
Further validation on the consumer dataset indicates that the model has a bias towards TV viewership, with higher bins watching more hours of television per week than lower bins (see Figure 6). Of note, however, is the drop-off in the model's very highest 20-tile; this may be the result of print and online media consumers "crowding out" the heavy TV watchers in this bin.

Figure 6. Average Hours of Television Watched in a Week by Individuals in Each 20-tile of the Media Consortium Look-Alike Model
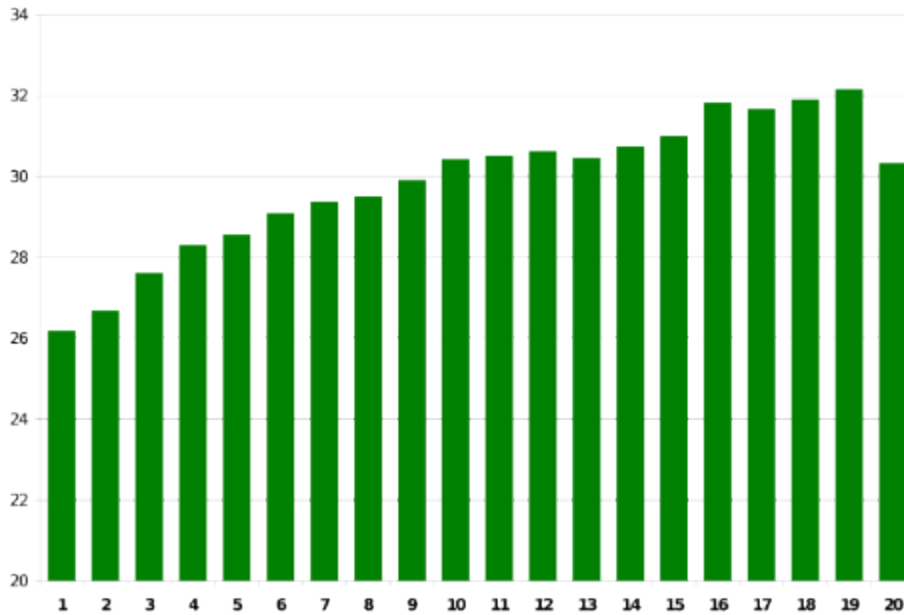
Modeling media consumption by channel

While the various media channels have differing distributions among the look-alike model scores, it may be of interest to The Media Consortium's members to have a more specific prediction of the primary channel of interest to each individual. Catalist attempted to find meaningful differences between the individuals affiliated with the various types of media organizations that are members of the consortium, with only limited success.

It appears that print and online publications are fairly correlated with a lot of overlap in their subscribers, so it was difficult to find characteristics that separated individuals that preferred one medium over the other. On the other hand, The Media Consortium includes members with mostly text-based communications (including both print and online channels) as well as members with an audio-visual format (including radio and television). We thought it may be promising to exploit this theoretical variation in communication channels, but again the model appeared to conflate various characteristics between the two groups, with no meaningful separation on external validation sets.

A model that predicts an individual's propensity to favor a particular type of media would need to be driven by The Media Consortium's specific needs. It seems possible to predict channel preference, but given the multitude of possibilities that could be explored it would make sense to guide the model building process based on a specific goal. Catalist is willing to provide further suggestions on this process if assistance is desired.

Conclusions

The look-alike model has been scored on all records in the Catalist database. While this document has shown charts based on 20-tiles of the model scores, the scores on the file will be percentiles of the model scores, thus allowing selection of more narrowly defined groups.

The greatest utility of a voter file model such as this one will likely be the pruning rented or internal prospect lists obtained by Media Consortium members. If no prospect lists are available, Catalist suggests combining the Media Consortium look-alike model with other voter file selects—such as single women, registered Democrats, or specific geographies—since the pool of very high scoring names is much larger and more diverse than a traditional rented or internal prospect list.

Further, the use of models to pull prospect universes from a database as large as Catalist's should always be seen as an iterative process, in which more is learned and the models perhaps are refined with each new wave of prospecting. Limiting target universes to only those individuals with the very highest model scores does not ultimately provide the learning opportunity that a wider selection would permit.

Overall, validation results indicate that the appropriate use of these models could identify prospecting universes that provide much higher response rates than a random sample of the voter file.

---

[1] The following constituent organizations submitted membership, subscriber, and/or donor data for this project: Afro-Netizen, Air America, Applied Research Center, Center for Independent Media, Feminsting.com, Free Speech TV, Grist TV, Grit TV with Laura Flanders, Hightower Lowdown, Link TV, Mother Jones, Ms. Magazine, One World, RH Reality Check, The American Prospect, The Nation, The Young Turks, and Yes Magazine.

[2] The models, data and information referenced herein were provided between the parties and are governed pursuant to the Data Evaluation Agreement between the Foundation for National Progress, and Catalist, LLC ("Catalist"), effective May 11, 2009. The information contained in this Memorandum is for informational purposes only. Accordingly, Catalist assumes no liability nor does it make any representations or warranties (express or implied) with respect to its condition, quality, completeness or accuracy.