Leveraging Data Mining and Geographic Information Systems to Gain Energy Efficiency Market Intelligence

Phillip Kelsven, Conservation Services Group

ABSTRACT

Market intelligence is an important key to running an effective residential energy efficiency program. Discovering which characteristics make up your typical customer profile and knowing where your highest potential customers are can make the difference in meeting your program goals. This paper will show how several commonly available data sources are used to perform a data mining analysis on your customer base and portray the information on informative maps.

Four data sources that are commonly available to energy efficiency program administrators are used to develop a data mining database. Data sources include utility customer consumption data, program participant data, local tax assessor data, and data from the US Census is used to gain intelligence about program participants and the larger potential market. Characteristics such as energy consumption, home size, home age, home value, household income, education, and age of home owner are some of the most informative data used in creating a profile. Data mining methods are successfully used in many consumer markets. Energy efficiency programs can also adopt these methods to maximize program participation and energy savings.

GIS mapping software is used to visually identify cities, neighborhoods, and households that are likely to be good candidates for energy efficiency improvements. Easy to use static or dynamic maps can then be utilized by program administrators to decide which areas to focus marketing efforts on. These methods are proven to increase customer response rates from direct mail and help identify neighborhoods to target for energy efficiency.

Introduction

Energy efficient products and services are like many other consumer products. There is a target market that is more likely to purchase energy efficient products and services than the general population. The unique differences in the target market are subtle, but when found they can be a useful program planning and marketing tool. Employing intelligence about who your likely customers are in the right way will increase customer investment in energy efficient products and services and increase participation in energy efficiency programs.

This paper will describe how Conservation Services Group (CSG) employs market research on a customer base in residential buildings to maximize customer participation and energy savings opportunities. Firstly, I will describe how commonly available data sources can be brought together to create customer profiles and segments which program marketing can be directed to. Secondly, I will describe how this data can be used to inform program planning by defining the potential market for energy efficiency in general, and for individual measures in particular. Thirdly, I will describe how Geographic Information Systems (GIS) can display this information on informative maps to help program staff develop strategies and tactics to enhance effectiveness. Finally, I will describe how all of this information is leveraged to employ a direct

targeted marketing campaign to utility customers who are the most likely to take advantage of your program offerings.

Data Sources

The activities described in this paper can be accomplished with data which is commonly available to utilities and energy efficiency program administrators at little to no cost. One of the most important data elements to have is utility customer energy consumption data. The data preferably needs to be the most current consumption data with monthly meter reads with the exact day, month, and year of reading. The consumption data should always be turned into a Normalized Annual Consumption (NAC) calculation by customer premise. Normalized Annual Consumption is a general term used to describe several particular techniques that normalizes energy consumption so that the consumption at any particular site at any time reflects the consumption at that site in a typical weather year. The NAC referred to in this paper is similar to the NAC routine employed in classical PRISM billing analysis (Fels 1986). The NAC not only accounts for differences in weather across the utility territory at different points in time, but also produces a base load, heating load, and cooling load estimation (for electric consumption) for each household. We typically find that customers who invest in energy efficiency have higher than average energy consumption levels, but this dynamic is found to differ by region or customer population. If a particular measure is being targeted, the base, heating, and cooling load estimations can identify potential customers who heat with electricity or have central air conditioning.

Program participant data is invaluable in creating a profile of your typical customer. The more data points you have of customers served, the more you will know about whom your customers are and what kinds of homes they typically live in. Customer profiling and segmentation will grow more robust every time you revisit the analysis with a bigger sample to analyze. A sample of at least a couple hundred participants is a good starting point, however a stronger profile can be built with thousands or tens of thousands of participants. There is a vast array of data points captured from program participants and the more data points captured the better. Program data can simply be used to identify customers, or can be more valuable by analyzing the particular characteristics of customers who install different types of measures. For example, customers who get incentives for energy efficient appliances may be significantly different than customers who do major weatherization improvements.

Local tax assessor data typically includes information on the physical characteristics of homes. Data on home square footage, volume, year built, construction type, and even presence of pools or hot tubs is extremely valuable information to have. It is common for weatherization programs to capture these characteristics for their participants, but the information becomes useful when compared to homes in the general population that do not participate in energy efficiency programs.

United States Census Bureau (US Census 2010) data is a rich source of information on certain housing characteristics and demographic data. If tax assessor data is not available, the median year built or proportions of homes built in ten year increments is available by census tract and block group. The breakdown of owner versus renter occupied housing units and single versus multi-family is available. Demographic indicators that are of value for market research available in the census data include median income and income by category, educational attainment, median age of homeowner, median home value, median household size, and many

more. The 2010 census does not have the level of accuracy that the 2000 census has due to elimination of the long form survey. The result of this is that data at the census block group level suffers from large margins of error, however accuracy is better at the census tract level and bigger geographic areas. For example, the most recent 2010 census often lists median income at the block group level with margins of error equal to or larger than the median income figure itself. The American Community Survey 2006-2010 (US Census 2010) data is the most recent and accurate at the census tract level which is useful for granularity in estimations.

There are many other sources of valuable data which vary in quantity and quality which can be useful. Some utilities have online home energy assessments which gather information from customers on their homes and ways they use energy. This data can be very useful in characterizing households that are interested in saving energy in their homes. There are several data mining companies that provide high quality data on individual households including home characteristics, demographics, and consumer spending patterns. Data from data mining companies can be very expensive, but is incredibly valuable for building profiles of your customers. Personal experience indicates that data from such companies can typically cost anywhere from 3 to 8 cents per record depending on what variables are selected and the size of the order. This cost can add up quickly for millions of records. Certain utilities and program administrators have employed aerial infrared cameras to capture the heat signature of building roofs (Cedar Falls Utilities 2010) This type of data can identify neighborhoods that appear to have many homes with significant heat leakage that may be good candidates for weatherization projects. Lastly, there has been a surge in online social networking sites such as Facebook where people voluntarily advertise their interests, behavior patterns, as well as social and political stances. The value of data from social networks for energy efficiency marketing remains largely unknown but could be extremely valuable. Facebook does not sell or provide data on their users to third parties but advertisers can market to a particular segment by describing the characteristics of that segment to Facebook. (Facebook 2012)

The dataset which will result from combining data sources will be based on your utility customer list with all of the additional data sources merged into the original customer list. It is inevitable that there will be missing data elements in the final dataset due to missing data or difficulties in merging in data by address or other unique identifiers. It must be noted that the census data is not at the household level and therefore the data is subject to bias, however some information is better than none. If care is taken to maximize data cleanliness and retain all available data, an incredibly informative dataset can be built.

Profiling and Segmentation

People and households who invest in energy efficiency tend to distinguish themselves in available data compared to households in the general population. Participants in energy efficiency financial incentive programs do not capture all households who invest in energy efficiency. We can only hope to profile and segment households who chose to participate in energy efficiency programs. Households who participate in energy efficiency programs may be different than households who invested in energy efficiency but did not take advantage of financial incentives to do so.

The differences in characteristics between program participants and non-participants will vary depending on market location, size of the market, as well as the particular product or service being offered. Program participants will be referred to as “participants” and those who do not

participate are called “non-participants. An example of differences among participants snd non-participants from internal CSG market research suggests that program participants in Oregon tend to have significantly higher incomes, larger and older homes, higher average gas and electric energy consumption, and tend to be more highly educated. Participants in Massachusetts do not display clear differences from non-participants for many indicators, and show opposite relationships for certain characteristics. There is not a one size fits all profile for households that participate in energy efficiency programs. The profiles change by geographic region and there are unique segments within each region’s participants that are doing energy efficiency for different reasons.

The goal in developing a customer profile is to describe your typical program participant with available data. A profile may emerge that tells a story about program participants. Table 1 below is actual data from a program that needs to remain unidentified that is still in its start-up phase. The table indicates that participants have greater average gas and electric consumption, live in older and larger homes, have higher incomes, greater home values, and tend to be older than non-participants. This information can be used to inform marketing messaging, develop a general geographic targeting strategy, or be used to market directly to customers that display similar criteria. A participant customer profile like this is a starting point that indicates participants tend to be different than non-participants as far as the data indicates. Getting to know your customers through data requires more in depth analysis.

Table 1. Example Participant vs. Non-Participant Profile

	Participants	**Non-Participants**
Average Annual Therms	1,414	1,258
Average Annual kWh	11,994	10,090
% pre-1980 homes	78%	72%
% Household income > $100k	42%	31%
Median Income	$90,234	$74,569
Median Home Value	$259,437	$194,695
Median Home Square Footage	2,095	1,936
Median Year Built	1961	1967
Median age of head of household	41	39

Source: Conservation Services Group internal report. 2012

Segmentation analysis is a technique used in marketing to identify homogeneous sub-populations within a market for a given product or service. Segmentation seeks to identify unique characteristics among these homogeneous populations that can be used to engage other people like them who may be interested in a particular product or service. Market segmentation is being utilized by energy efficiency programs to maximize customer participation. A good discussion of segmentation applied to energy efficiency appears in Moss and Cubed (Moss & Cubed 2008).

Customer segmentation analysis assumes that your program participants are not the same and typically display significantly different characteristics. Customer segmentation can be done with easily acquired data, but becomes increasingly useful if more detailed data on participants and non-participants is gathered through surveys or data mining companies. Segmentation analysis uses as many common data points as possible to create unique segments within your

customer base. A statistical clustering technique is often used to create segments which must be meticulously analyzed for trends and differences in relation to other segments. Statistical clustering techniques are employed to help build homogeneous groups of observations that have many different variables. Grouping observations manually with many variables is time consuming and imprecise, where computing power can recognize patterns in data that are extremely difficult to detect by manual methods. Detailed household level data from data mining companies will allow such things as household makeup, number of children, age and occupation of occupants to be known as well as spending patterns and political affiliations. Surveys of participants and non-participants about their reasons for doing or not doing energy efficiency are the most valuable information but require samples numbering in the hundreds which can be expensive to conduct. Field staff who are in peoples' homes everyday are often valuable sources of anecdotal information and may be able to offer insight into segmentation analysis.

A chosen number of unique segments are developed for further analysis. Names are given to the segments which seem to characterize a majority of the segment or describe its distinguishing characteristics. An example of segments created in a segmentation analysis appears in table 2. Often a large share of each segment will display common characteristics, but a distinguishing characteristic makes itself known by careful analysis. For example the "Retired on a Budget" segment is named so because a majority of this group is found to be retired and has a median income that is lower than the typical program participant. The "Power Family" acquired that name because a majority of the segment are families with children who have a significantly higher income than the typical program participant. The results of a segmentation analysis are best used to develop specialized marketing messages which are likely to resonate with the respective segments. At minimum, a better idea of who most of the program participants are can inform standard marketing messaging.

Table 2. Example of Segments

Segment	% of Customers
Retired on a Budget	32.8%
Making Ends Meet	26.4%
Comfortable Empty Nesters	25.6%
Average Joes	9.2%
Power Family	6%

Source: Conservation Services Group internal report. 2012

Segmentation on a budget can be accomplished with the 4 four commonly available data sources; consumption data, census data, tax assessor data, and program data. Segments are created that go beyond the initial participant non-participant profile to identify several unique customer groups. To avoid bias from census data which is at the census tract or block group level, the consumption data, program data, and tax assessor data can all be aggregated up to the census tract or block group level. The census tracts or block groups become the level of analysis and are segmented into unique groups. Granularity is lost in doing this but if the goal is to identify neighborhoods or towns to target, this can be a valuable exercise.

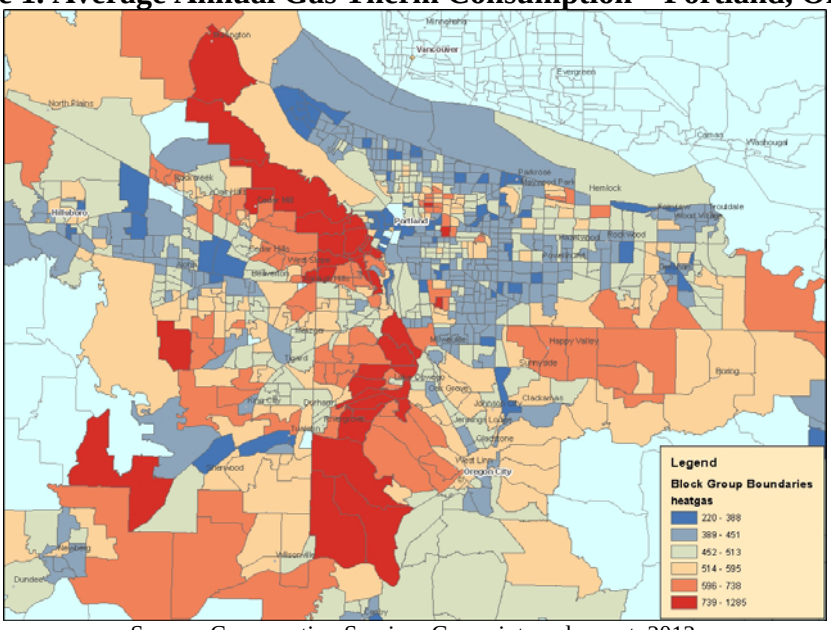
Mapping Data on Geographic Information Systems

Participants live in houses which are spread out across utility territories. All data on customers can be analyzed geographically on maps and data elements can be displayed simultaneously in multiple layers. Geographic patterns in the data emerge which could never be explicitly known without displaying the information geographically. GIS software is user friendly but requires a base level of training to fully understand how to manipulate data with geographic features. A determined beginner who knows how to handle data can master basic GIS concepts and operations in a short amount of time.

Geographical representation of data displays known relationships in the data and can uncover surprising trends. Mapping energy consumption, home square footage, and income verifies established knowledge that all of those elements are interrelated. An example of this is displayed in Figure 1 which shows the annual therm gas consumption in Portland, Oregon in 2009[1]. The warmer colors in red and orange are census blocks that have the highest energy consumption by sextiles. The long vertical red strip on the map represents homes with some of the highest average gas energy consumption in Oregon. Maps of home square footage and income also isolate this area as some of the largest homes with the highest median household incomes in Oregon. Figure 2 shows the average energy intensity (therms/home square footage) in Portland, Oregon by census block group. This map indicates that pockets of homes in East Portland have high average energy intensities and may be good potential areas for weatherization. East Portland has some of the oldest homes in Oregon, which also tend to be smaller than average. Energy intensity is an imperfect measure of energy using efficiency since smaller homes are penalized by energy intensity calculated using square footage. Home volume is a better measure to use which can be calculated with some assumptions about ceiling height but is not data that is often available on a large scale. These maps are good examples of market intelligence that is valuable for program planners to develop strategies and tactics for marketing and outreach.

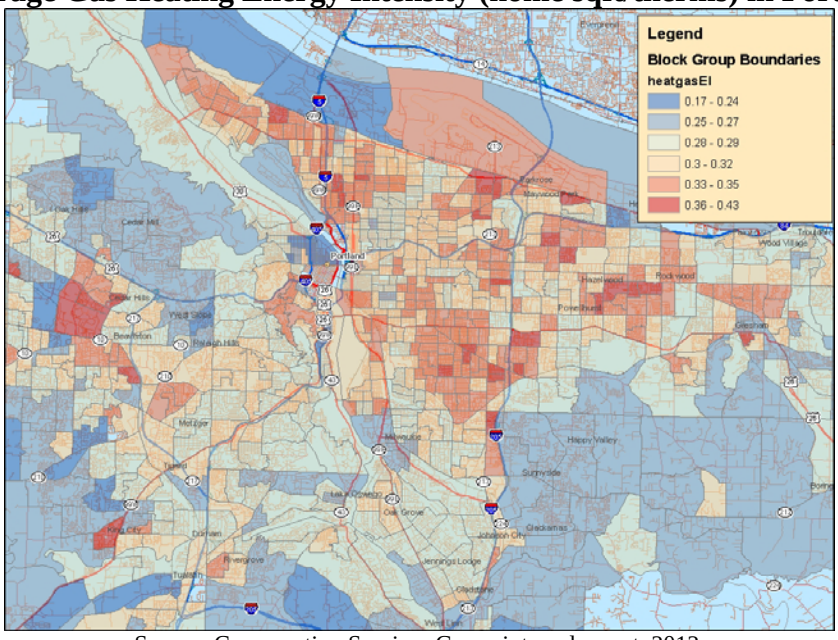
[1] Data provided Northwest Natural Gas and Energy Trust of Oregon. Gas consumption is Normalized Annual Consumption (NAC) in 2009

Figure 1. Average Annual Gas Therm Consumption – Portland, Oregon



Source: Conservation Services Group internal report. 2012

Figure 2. Average Gas Heating Energy Intensity (home sqft/therms) in Portland, Oregon



Source: Conservation Services Group internal report. 2012

What to map and how to map it are skills learned after representing many data elements in map layers in different geographic scales. Making maps that are informative requires mapping the right elements at the right level of granularity. Mapping of energy consumption, energy intensity, housing characteristics such as square footage and year built, program participation, and demographics such as income tend to be the most practical and informative. Representing the average or median values by census block group tend to produce maps with the necessary granularity without taking away from analysis of the big picture. Census block group

identifiers known as "FIPS"[2] are easily merged into data sets by using geocoding tools in GIS software. Producing dots on a map representing households is often un-informative unless you are looking at the neighborhood or street level. Representing data at very large geographic scales can also be uninformative particularly in dense geographic areas where the census block groups are very small. Program staff often want to see data by zip codes since that is a boundary which everyone is familiar with and is readily available in customer data. Analysis at the census block group level should be encouraged since it provides an increased level of detail.

There are many practical questions that can be answered with GIS. I often get requests to display utility territories by zip code, find out how many electrically heated homes are in a certain county or town, or find out which homes in a list of addresses are within an official city boundary. Requests like these often do not require GIS since the information can be gathered directly from the data itself. However, preparing maps with this information can cut down on the number of requests if the data is readily available in prepared maps for program staff to access at any time. Numerous static maps in jpeg or tif formats can serve this purpose, however developing interactive maps with this information is much more useful.

Interactive maps which can be manipulated by novice GIS users are incredibly useful and efficient. ESRI, who produces the most popular GIS software has developed a free GIS tool called ARC Explorer which allows people without GIS experience or expensive software to download layer files produced by more experienced GIS users. ARC Explorer allows many layers to be brought into an interactive map which allows for panning and easy manipulation of map scale. City and street layers add perspective to the maps which allows precise targeting for program planners and field staff. Certain data such as point data is amenable to Google Maps, however Google Maps does not handle representation by boundaries such as block groups. Experience with ARC GIS Explorer has shown that the software requires a lot of RAM and processing power. Improved GIS tools for easier access will make analysis by novice users easier.

GIS analysis and mapping are tools are not widely used among energy efficiency program administrators. An increasing number of program administrators, implementers, and consulting firms are realizing the benefits that GIS brings to the table including analysis of housing stocks, program activity, and demographic research. There are an incredible amount of experienced GIS users across the country who are working in other industries and sectors that could offer their skills to enhance the market research functions for energy efficiency programs.

Program Planning

This paper has addressed the data sources, tools, and methods available to analyze data for greater energy efficiency market intelligence. Now we will address the practical use of the information. Program administrators and implementers are increasingly facing aggressive savings and participation goals. More informed market research will help set realistic goals based upon the potential market and help to craft targeting strategies to maximize participation, installation of major measures, and energy savings.

Knowing the size and location of your potential market are key drivers to goal setting and program expectations. The data that is culled from our four commonly available sources can go

[2] Federal Information Processing Standards – "A standardized set of numeric or alphabetic codes issued by the National Institute of Standards and Technology (NIST) to ensure uniform identification of geographic entities through all federal government agencies." U.S. Census Bureau- http://quickfacts.census.gov/qfd/meta/long_fips.htm

a long way to aid informed decision making for program planning. Tax assessor and census data on housing characteristics can estimate the potential market for weatherization based upon home vintages, housing type, and even date of remodels from assessor data. Combining this with knowledge of historical building codes can inform which home vintages are most likely to need insulation. The potential market for new HVAC or domestic hot water equipment can be estimated by knowing the effective useful lives (EUL) of the equipment and the vintages of homes. For example, homes that are 20 years old are due for new gas furnaces or homes that are 15 years old are due for new heat pumps. The number of older homes due for new HVAC equipment can be extrapolated by assuming that homes that are 40, 60, and 80 years old for example are due for new gas furnaces. Yet another example is estimating the market for efficient gas hearths to homes that are likely to have fireplaces that can be converted, or estimating the market for solar hot water to homes that are known to have pools. Many market research questions can be answered with the data that is compiled.

Development of energy load profiles and potential savings in different housing types is yet another use of the data. Customer consumption data that is appended with tax assessor and prior participant data can be used to estimate average energy consumption in different housing types. For example energy use load profiles in multifamily, small multifamily, and mobile homes can be estimated with available data. This can inform not only the potential market but the energy savings that can be expected from installing measures at these sites. Your program may be missing out on savings or over estimating savings due to improper estimations of the typical pre-treatment energy consumption. Similarly, the number of large, average, and small single family homes and the savings that can be expected from them are potential uses of the data. If actual verified savings from impact evaluations can be matched to the treated homes it can provide an excellent data set from which expected savings based on installed quantities, home size, and age can be estimated and used for program planning.

Strategic planning may require localized goals which can be intelligently created with quality data. Maximization of site energy savings can be accomplished by targeting areas with a lot of older large homes with above average energy consumption. If projected savings are modeled based on home size or calibrated on pre-treatment consumption, treating large homes can help achieve goals faster. Additionally, homes with above average energy consumption are often found to have higher savings realization rates in impact evaluations which will improve program effectiveness over time. Analysis of program participant data may produce findings that certain areas or neighborhoods are more likely than others to install major measures or are targets for a particular type of measure. These examples only name a few of the myriad of uses of our data for use in program planning.

Targeted Marketing

Predictive analytics is a growing field of marketing using statistics and market research to market to households and people that are likely to purchase a given product. Many industries are taking advantage of this growing field of statistics to significantly increase sales and profits. Industries as varied as insurance, baby products, finance, and journalism are using predictive analytics to their advantage. The field has also expanded into sports and politics where undervalued players are identified by their statistics (Lewis 2003) and probable voters of certain political persuasions are identified by available data (Sifry 2011) The energy efficiency sector

would be smart to adopt predictive analytics in a major way to meet aggressive and necessary goals.

Marketing in the energy efficiency sector has historically been about mass marketing campaigns. Energy efficiency programs have relied on mass direct mail, utility bill inserts, as well as tv and radio ads. These types of marketing campaigns work well for general program awareness, but fail to relate to customers directly. Marketing dollars could be spent more intelligently by reaching out to utility customers that appear to be good candidates for energy efficiency. For example, a large direct mail marketing campaign can expect a 1/2 to 1 percent response rate of households who end up installing measures. In CSG's experience, response rates can easily be doubled or tripled to rates approaching 3 to 4 percent by employing predictive analytics to choose a smaller group of utility customers to mail to that are likely to be more responsive to marketing (CSG 2012)

CSG has used targeted marketing strategies for many years to run effective energy efficiency programs. More recently CSG began using predictive analytics and data mining to gather more data on participants and potential participants to make targeted marketing more informed and effective. CSG uses econometric and neural network models to produce probabilities that a non-participant will be likely to install energy efficiency measures and take advantage of financial incentives. CSG then markets directly to utility customers that have the highest probabilities of participation. The amount of data available on utility customers often varies which poses challenges for robust predictive analytics. Key data pieces such as consumption data or tax assessor data are not always available to build a participant profile. Even in those cases, some data is better than none and we use what we have to target likely participants. Census data is always available and can at minimum inform which areas are going to be the most likely to have potential participants.

CSG has evidence that targeted marketing produces better response rates than random non-targeted direct marketing. CSG tested our targeted marketing methods in a pilot whose details cannot be released yet. The pilot mailed postcards to three separate groups at the same time. A primary group consisting of the top quartile probability households was mailed to as well as secondary group consisting of second quartile probabilities. A third group of randomly selected households were also mailed to as a control group. The primary group resulted in significantly better response rates of over 5.3%, where the secondary and control group had response rates of 3-4%. All of these response rates are considered very good compared to the industry average of .5% – 1%. The program is in a mature energy efficiency market with generous incentives. 1% may not sound like much of a response rate premium, however when the number of customers mailed to is scaled into the hundreds of thousands, the increase in participants can be significant and may be the difference in meeting program goals. The participants gained are also likely to have greater savings since in many markets they tend to live in larger homes with above average energy consumption. Targeted marketing using predictive analytics works.

The most effective predictive analytics is built from large samples of households that have installed energy efficiency measures in their homes. The larger the sample of participants, the more confident we are that the participant profile is an accurate portrayal of our typical customer. As time goes by and programs gain more participants we revisit the participant profile and update it with the new sample of participants. The predictive models used are a form of artificial intelligence, meaning that the model has the capability to continuously learn as new data is fed into the model.

Targeted marketing is often most needed in programs that are in their start-up phase. Energy efficiency programs that are new to a region have the difficult task of raising program awareness. Targeted marketing is an effective way to begin to get participants into the program. Start up programs pose a challenge from a predictive analytics perspective. There are no prior program participants to build a profile of or use to inform a model. In these cases we use what we know about participants in other regions with similar programs. For example if the new program is providing free home energy assessments with the aim of converting households into major measures we use profiles of participants in a similar program in a region with similar climate, housing stock, economic characteristics, and population. At minimum the first round of targeted marketing could be based solely on home characteristics that meet certain criteria such as built before 1980 and over 2,000 square feet. As soon as a couple hundred households have participated in the program an informed targeted marketing strategy can be developed.

If targeting customers directly is not a program tactic, a more general geographic targeting strategy can be developed. All available data can be aggregated up the census block group, census tract, zip code or county level. Areas that are shown to have high proportions of participants can be analyzed for distinguishing characteristics. The chosen geographic boundary can be treated just like you would treat individual participants in developing a participant profile. The main difference is that an emphasis is placed on calculating proportions such as the proportion of homes that are over 2,000 square feet, or the proportion of homes above a certain income level. Areas that have high numbers of participants are considered to be the participant profile and then the targeted marketing strategy is to market to areas that display similar characteristics. A segmentation analysis can even be conducted on the chosen geographic areas. For example a clustering analysis on census block groups can be conducted and segments developed and named for their distinguishing characteristics.

The effectiveness of targeted marketing using predictive analytics can vary depending on the maturity of the market for energy efficiency. Anecdotally, CSG has noticed that direct mail targeted marketing seems to produce a better response rate in mature markets where programs are long standing and households have a base awareness of the program and may be familiar with its offerings. New markets for energy efficiency programs can be difficult to penetrate until a base level of program awareness is established through other marketing channels or word of mouth. For direct mail, the message itself, the timing of the messaging, and the level of incentives can also be incredibly important factors in the success of a targeted marketing effort. There are several potential reasons for a poor response rate, and the reasons can be hard to identify. Often times trial and error is necessary to learn what works well.

Conclusion

This paper presented a methodology to gain market intelligence for energy efficiency programs. Four commonly available data sources are all that is needed to begin analysis which can provide a rich source of information on program participants and the greater potential market. A methodology for developing a participant profile and possible segmentation analysis describes possible uses of the data. Use of GIS mapping technology to gain an understanding of geographical dispersion and where potential participants may live is a useful aid. This data can be used to improve marketing and market intelligence, as well as inform program planning. Predictive analytics is one of the most valuable tactics that the data can be used for. Now it is time to get to work and make your program more effective!

References

Cedar Falls Utilities. 2011. "Spy in the Sky: Aerial Thermal Imaging in the Digital Age" http://www.iowaenergy.org/Portals/0/2011%20Summit/Presentations/Thermal%20Flyover.pdf

Conservation Services Group. 2012. Internal un-published and confidential report. Portland, Oregon

Facebook. 2012. Data Use Policy. https://www.facebook.com/about/privacy/ . Menlo Park, California.

Fels, Margaret F. (1986) "PRISM: An Introduction." *Energy and Buildings*, (9) p.5-18.

Moss, Steven and Cubed, M. (2008) "Market Segmentation and Energy Efficiency Program Design" Prepared for California Institute for Energy and Environment.

Lewis, Michael. 2003. "Moneyball: The Art of Winning an Unfair Game."W.W. Norton & Company.

Sifry, Micah. "How Obama's Data Crunching Prowess May Get Him Re-elected." CNN-Tech. http://www.cnn.com/2011/10/09/tech/innovation/obama-data-crunching-election/index.html

U.S. Bureau of the Census. 2012. www.census.gov. Washington, D.C.: U.S. Bureau of the Census.

U.S. Bureau of the Census. 2010. American Community Survey. http://www.census.gov/acs/www/. Washington, D.C.: U.S. Bureau of the Census