Energy Efficiency

Exhibit 7
Supplement 1
Residential Propensity Modeling Study

January 1, 2013

PUGET SOUND ENERGY
The Energy To Do Great Things

Vertex Professional Services

Energy Efficiency Modeling and Pilot

PUGET SOUND ENERGY

Propensity Modeling
Model Build and Validation
Results Readout

Oct 3rd, 2012

VERTEX BUSINESS SERVICES

Contents

Project Background & Status

Executive Summary

Completed Work:

6 Predictive models built and validated

- Models focus only on single family existing homes
- **>984,000** Accounts have received **6** predictive scores
- Each model has an average of **19** variables
- Predictive models determined to be **220% - 431%** more effective at identifying energy efficiency program participants than random selection

Next Steps:

4 Campaigns identified for pilot

- Determined based upon conversations with program and market managers
  - *Water Heat*: Go-live October 22nd
  - *Weatherization:* Go-live early Nov
  - *Refrigerator Decommissioning:* Go-live mid Nov
  - *HomePrint:* Go-live in early Dec

Campaign response will be measured for each campaign

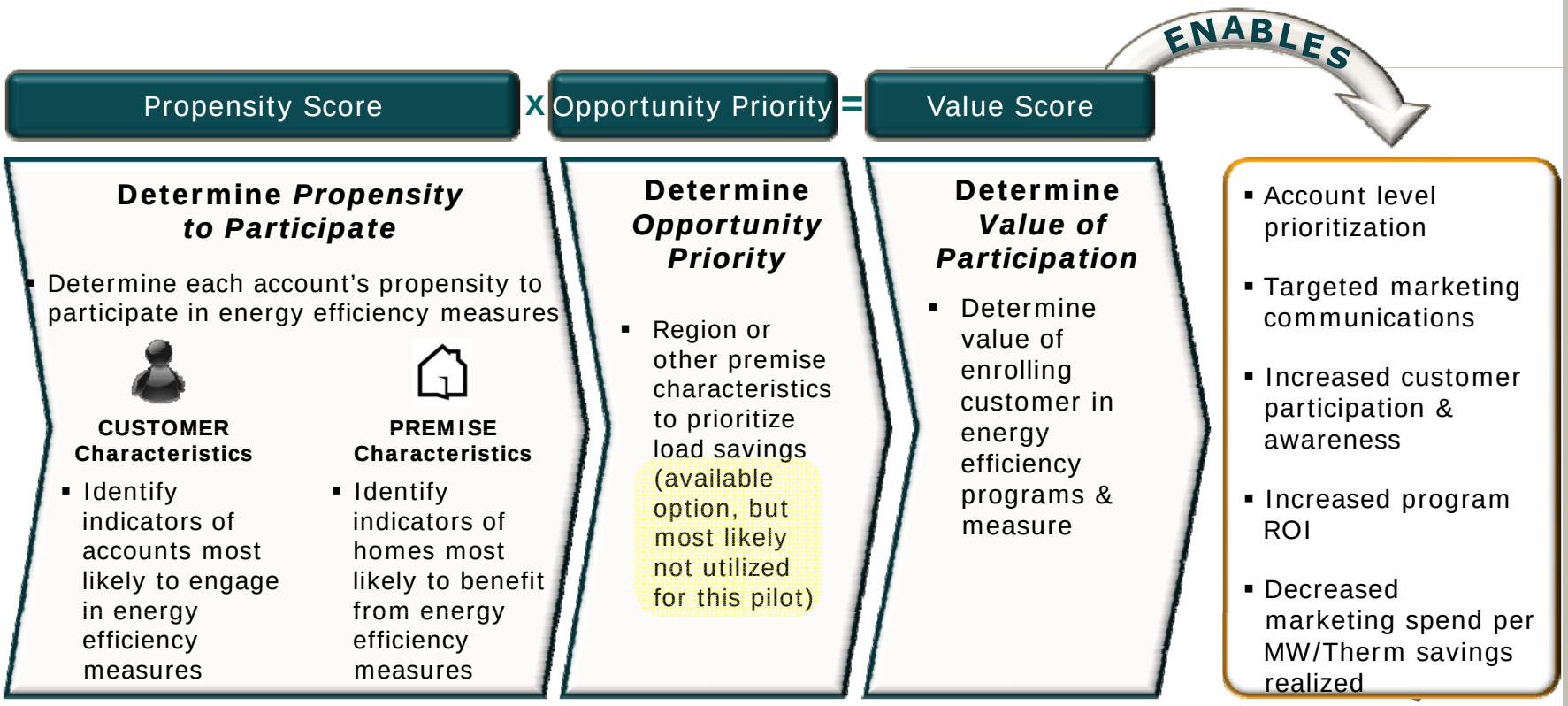
- Metrics to include *Rebate Processing Rate & Rebate Submittal Rate*
- Will also measure *Customer Inquiry Rate* where possible
- We will deliver the measured results readout in mid-late February
- We will have checkpoint discussions prior to this readout

Project Background & Status

PSE Predictive Modeling – Basic Approach

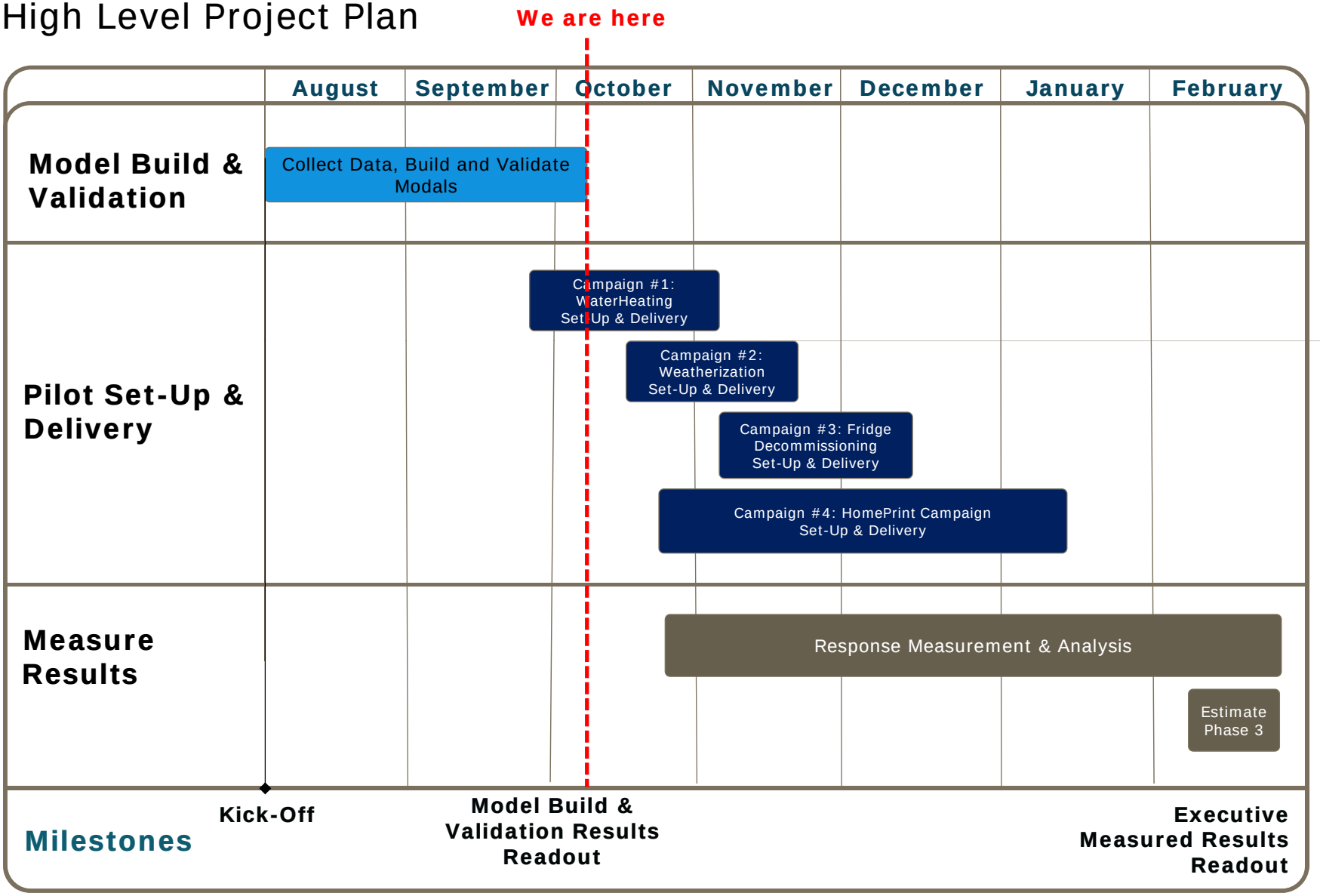
Predictive modeling is a customer focused, data driven and analytics based approach that produces account level prioritization, which enables targeted marketing of energy efficiency measures to the right customers.

The proposed solution is system and technology agnostic.



Project Background & Status

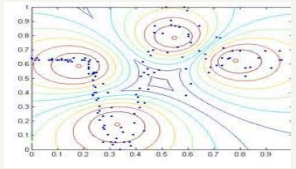
High Level Project Plan



Contents

Approach

Methodology to Produce PSE Customer Scores

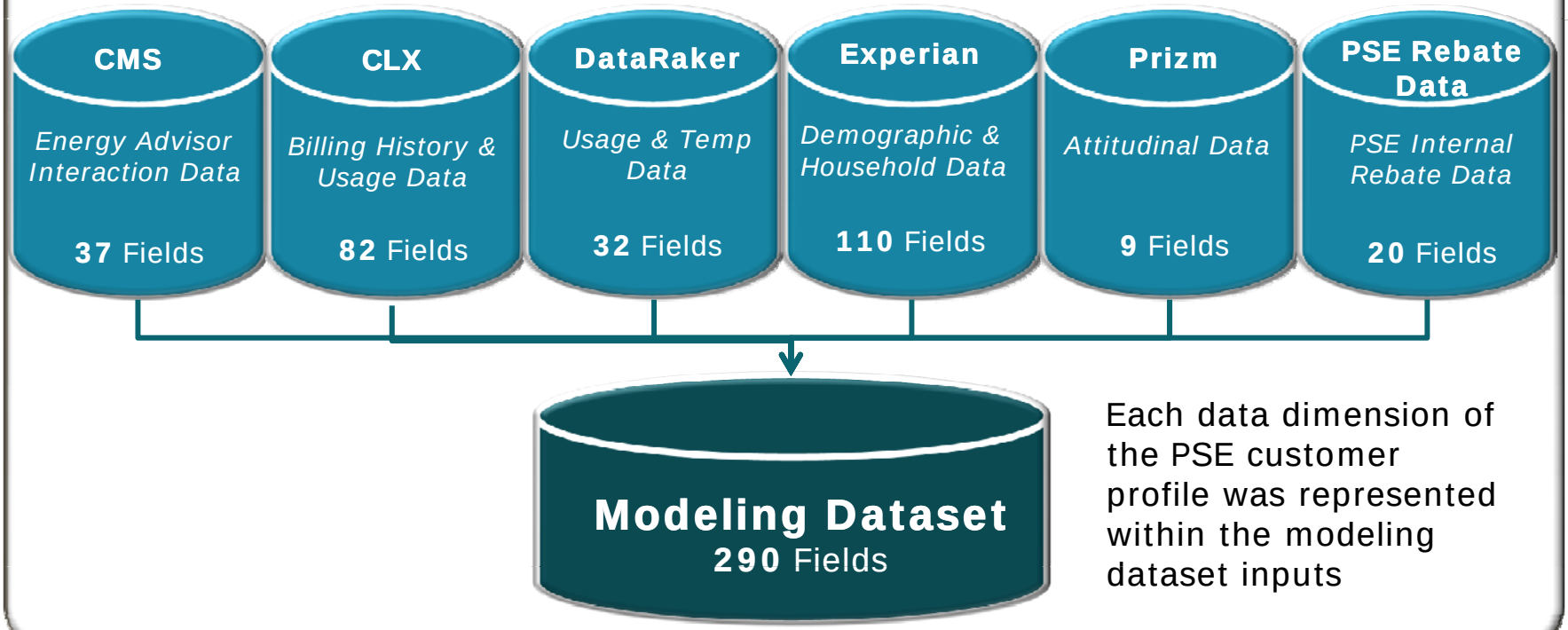
Create Groups	Build Models	Validate Models	Produce Scores
Completed 9/6/2012	Completed 9/23/2012	Completed 9/25/2012	Completed 9/27/2012
Approach: ▪ Cluster analysis performed to group 14 EE programs into cluster sets based on similarities in purchase behaviors	**Approach:** ▪ Build one predictive model per program grouping (5 models) ▪ Build one general propensity to participate model	**Approach**: ▪ Test models for predictive accuracy against hold out samples	**Approach:** ▪ Execute models to score each account on likelihood to participate $f(z) = \frac{1}{1+e^{-z}}$
Product: ▪ 6 dependent variables identified 	**Product:** ▪ 5 predictive models (one per program grouping) ▪ 1 general propensity to participate model	**Product:** ▪ Validation of model accuracies	**Product:** ▪ 5 program group propensity scores per acct. ▪ 1 general propensity to participate score per acct. ▪ >984,000 residential existing accounts scored on 6 propensities
Use: ▪ Dependent variables for propensity models	**Use:** ▪ Predict likelihood of customers to participate in program ▪ Predict general propensity to participate	**Use:** ▪ Confirms strength of prediction for each model	**Use:** ▪ Tool to enable targeted marketing outreach

Approach

Data Scope

- Total number of fields evaluated for possible use in models: **290**
- PSE sourced data was well populated except for DataRaker which was not used in the model
- Match rate across datasets was **97%**
- Total number of variables determined to be predictive: **69**

Data evaluated for use in models:



Each data dimension of the PSE customer profile was represented within the modeling dataset inputs

Approach

Program Clusters & Final Models

A cluster analysis was completed to group the rebate programs into a **manageable number of natural sets.** The results of the cluster analysis informed the final program groupings for the model builds.

The 6 program groupings outlined below represent the final 6 models that were built, validated, and used to provide propensity scores to all active, single family, residential accounts.

Final Predictive Models

1 Appliances

- Refrigerator Decommissioning
- Energy Star Appliances – Refrigerators
- Freezer Decommissioning
- Energy Star Appliances – Clothes Washers

2 Heating

- Single Family Water Heat
- Single Family Space Heat

3 Weatherization

- Single Family Weatherization
- Single Family Weatherization – Windows

4 Refrigerator Replacement

- Refrigerator Replacement

5 HomePrint™

- HomePrint™

6 Any / All Programs

- Overall Propensity to Participate

Contents

Results

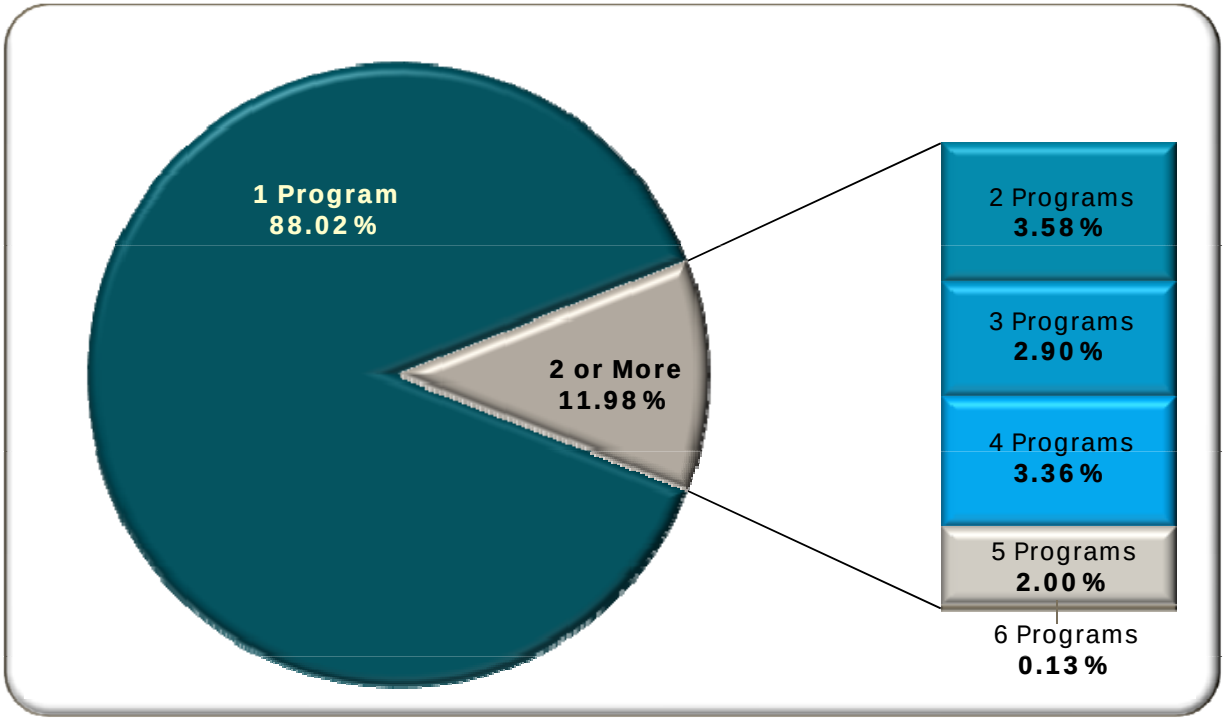
Value of Targeted Marketing

Model Exclusivity:

Minimal overlap was found within the high propensity populations across models.

- **234,149** unique accounts representing 24% of total modeled accounts were analyzed to determine overlap
- Only **11.98%** of accounts were highly likely to participate in more than one program.

High Propensity Account Overlap Across Models



- Less than **12%** of accounts identified as highly likely to enroll in 1 program were highly likely to enroll in additional programs
- We will account for multiple campaign exposures when selecting champion/challenger lists for each campaign

Results

Interesting Relationships

High propensity to enroll in any program accounts tend to:

- Be affluent, with higher incomes
- Have good credit and billing histories
- Have discussed one or more programs with Energy Advisors
- Have higher monthly bills and usages
- Be homeowners

Any interaction with an energy advisor:

- Commonly appears as a strong predictor of a customer‘s propensity to enroll

Refrigerator Replacement model:

- Appears to have a dramatically different customer profile than other models
- Represented by typically lower income customers

Additional notes about variables:

- Weather region appears to be a commonly occurring predictive variable
- Attitudinal profile appears to have an affect on a customers propensity to enroll
- Above average usage and bill amount appear to be good predictors of a account’s propensity to participate in a Weatherization program

Results

Interesting Relationships

16X

Customers who contact an Energy Advisor and discuss HomePrint™ are 16 times more likely to participate in an Energy Efficiency program.

8X

Customers who discuss one or more products with an Energy Advisor are 8 times more likely to participate in a Weatherization program.

3X

Customers with high monthly utility usage are 3 times more likely to participate in the Refrigerator Replacement program.

2X

Customers with high monthly utility usage are 2 times more likely to participate in an Appliances program.

Results

Model #1: Appliances Model

A stepwise logistic regression was conducted to identify **18 significantly predictive variables** for customers generally likely to enroll in any Appliance rebate program.

	Variable Description	Source	Wald Chi Square
1	Account holder is a homeowner	Experian	**710.03**
2	Secondary electric service account - low voltage, typical for residential	CLX	**391.54**
3	Total number of products discussed with EA between Jan 2010 - June 2012	CMS	**353.25**
4	Account received a late payment fee, from delinquency history	CLX	**339.65**
5	Regulatory asset tracker credit applied to account	CLX	**133.02**
6	Account located in the Bellingham weather region	CLX	**112.52**
7	Account holder extremely likely to be married	Experian	**93.00**
8	Average usage quantity >= 1000 kWh	CLX	**52.03**
9	Gas service account	CLX	**51.72**
10	Average usage quantity 150-999 kWh	CLX	**47.65**
11	Account is on budget billing plan	CLX	**47.60**
12	Household located in region with a certain median years of education	Experian	**35.35**
13	Account located in the Tacoma weather region	CLX	**28.25**
14	Average monthly bill	CLX	**23.97**
15	Household located in region with a certain median age	Experian	**23.78**
16	Count of children within household under 18 years of age	Experian	**18.55**
17	Account located in region classified as "Money Brains": city dwellers, high incomes, advanced degrees, sophisticated tastes	Prizm	**18.26**
18	Account located in region classified as "American Dreams": ethnically diverse, multilingual, middle-class	Prizm	**11.09**

The comparative size of the Wald Chi Square statistic is a rough measure of the **importance of the variable** in predicting responses

220% more predictive than random selection

44% of participants can be found within **20%** of the population

Results

Model #2: Heating Model

A stepwise logistic regression was conducted to identify **26 significantly predictive variables** for customers generally likely to enroll in any Heating program.

	Variable Description	Source	Wald Chi Square
1	Total number of products discussed with EA between Jan 2010 - June 2012	CMS	**537.17**
2	Heating & water heating rebates discussed w/ EA	CMS	**349.22**
3	Gas service account	CLX	**184.70**
4	Household located in region with a certain median years of education	Experian	**166.33**
5	Account holder is a homeowner	Experian	**136.43**
6	Average usage quantity >= 1000 kWh	CLX	**130.96**
7	Account holder's length of residence	Experian	**108.07**
8	Account located in the Olympia weather region	CLX	**75.33**
9	Historical number of credit events	CLX	**42.66**
10	Household located in region likely comprised of individuals with blue collar occupations	Experian	**39.03**
11	Account holder's age	Experian	**37.72**
12	Customer received rebate flyer brochure from EA	CMS	**37.47**
13	Account located in the Burien weather region	CLX	**31.56**
14	Showerheads discussed with EA	CMS	**30.23**
15	Household located in region classified as "Metropolitan Struggler": low-income, urban household	Experian	**24.62**
16	Account located in region classified as "Blue Blood Estates": wealthy, suburban, luxury lifestyle	Prizm	**22.89**
17	Account received a miscellaneous gas service charge	CLX	**21.90**
18	Account located in region classified as "Big Fish Small Pond": older, upper-class, college-educated, professionals	Prizm	**14.24**
19	Average monthly bill	CLX	**13.33**
20	Was a credit event point charged for a final notice, account was not sent notice	CLX	**12.67**
21	Was a disconnect order scheduled, account was disconnected	CLX	**12.40**
22	Household located in region with a certain median age	Experian	**11.51**
23	Account located in the Bellevue weather region	CLX	**10.38**
24	Account located in region classified as "American Dreams": ethnically diverse, multi-lingual, middle-class	Prizm	**8.11**
25	Account located in region classified as "Fast-Track Families": upper middle-class, children present, disposable incomes	Prizm	**6.17**
26	Account located in region classified as "Accumulated Wealth": wealthy, upscale, suburban	Prizm	**6.06**

251% more predictive than random selection

50% of participants can be found within **20%** of the population

Results

Model #3: Weatherization Model

A stepwise logistic regression was conducted to identify **19 significantly predictive variables** for customers generally likely to enroll in any Weatherization program.

	Variable Description	Source	Wald Chi Square
1	Account located in the Bellevue weather region	CLX	**701.79**
2	Total number of products discussed with EA between Jan 2010 – June 2012	CMS	**304.64**
3	Account located in region classified as "Gray Power": aging, middle-class, homeowners, suburbanites	Prizm	**246.04**
4	Estimated current home value	Experian	**220.85**
5	Was a credit event point charged for an urgent notice, account was not sent notice	CLX	**176.48**
6	Account located in region classified as "New Empty Nests": active, older, raised children recently out of the house	Prizm	**140.44**
7	Average usage quantity 150-999 kWh	CLX	**102.80**
8	Contractor referral service discussed with EA	CMS	**80.32**
9	Average usage quantity >= 1000 kWh	CLX	**74.98**
10	Household located in region classified as “Bourgeois Prosperity”: high-income, suburban households	Experian	**72.40**
11	Account located in region classified as "Money Brains": city dwellers, high incomes, advanced degrees, sophisticated tastes	Prizm	**62.97**
12	HomePrint™ discussed with EA	CMS	**47.59**
13	Heating & water heating rebates discussed w/ EA	CMS	**47.17**
14	Account located in region classified as "Young Digerati": tech-savvy, singles and couples, fashionable neighborhoods	Prizm	**45.53**
15	Account located in region classified as "American Dreams": ethnically diverse, multilingual, middle-class	Prizm	**37.98**
16	Household located in a region with a certain % population above 18 years of age	Experian	**34.36**
17	Account located in the Tacoma weather region	CLX	**28.50**
18	Household has kids present	Experian	**21.08**
19	Household located in region with a certain median age	Experian	**13.43**

293% more predictive than random selection

59% of participants can be found within **20%** of the population

Results

Model #4: Refrigerator Replacement Model

A stepwise logistic regression was conducted to identify **16 significantly predictive variables** for customers generally likely to enroll in a Refrigerator replacement program.

	Variable Description	Source	Wald Chi Square
1	Gas service account	CLX	**82.19**
2	Was a credit event point charged for a disconnect order, account was not disconnected	CLX	**28.35**
3	Average monthly bill	CLX	**23.88**
4	Account received a late payment fee, from billing history	CLX	**17.25**
5	Account holder is a female	Experian	**15.14**
6	Account holder extremely likely to be married	Experian	**13.90**
7	Highest historical amount past due during any delinquency period	CLX	**13.32**
8	Total number of products discussed with EA between Jan 2010 - June 2012	CMS	**11.48**
9	Estimated current home value	Experian	**10.72**
10	Household located in region likely comprised of individuals with blue collar occupations	Experian	**7.33**
11	Account holder prefers information by direct mail	Experian	**6.94**
12	Account located in region classified as "Sustaining Families": economically challenged, families, working to make ends meet	Prizm	**5.11**
13	Was a disconnect order scheduled, account was disconnected	CLX	**4.73**
14	Account located in region classified as "Striving Singles": working, single	Prizm	**4.65**
15	Final delinquency notice was sent	CLX	**3.26**
16	Account located in region classified as "Old Milltowns": old mining/manufacturing towns, retired, downscale incomes	Prizm	**3.22**

431% more predictive than random selection

86% of participants can be found within **20%** of the population

Results

Model #5: HomePrint™ Model

A stepwise logistic regression was conducted to identify **16 significantly predictive variables** for customers generally likely to enroll in any HomePrint™ program.

	Variable Description	Source	Wald Chi Square
1	HomePrint™ discussed with EA	CMS	**6216.66**
2	Household located in region with a certain median years of education	Experian	**573.89**
3	Gas service account	CLX	**555.47**
4	Number of unique services - gas and/or electric	CLX	**372.64**
5	Account located in the Burien weather region	CLX	**296.91**
6	Account located in the South Seattle weather region	CLX	**244.11**
7	Account located in the Ellensburg weather region	CLX	**191.80**
8	Married couple household	Experian	**171.73**
9	Green Power discussed with EA	CMS	**145.77**
10	Account located in region classified as "Big Fish Small Pond": older, upper-class, college-educated, professionals	Prizm	**129.32**
11	Contractor referral service discussed with EA	CMS	**109.87**
12	Account located in region classified as "Country Squires": wealthy, ex-urban residents living small-town lifestyle	Prizm	**100.95**
13	Account located in region classified as "God's Country": upper-income couples in spacious homes	Prizm	**90.55**
14	Account located in region classified as "Sunset City Blues": retired, lower-middle-class, singles and couples	Prizm	**17.13**
15	Household located in region likely comprised of individuals with blue collar occupations	Experian	**13.32**
16	Estimated current home value	Experian	**9.24**

402% more predictive than random selection

80% of participants can be found within **20%** of the population

* The fact that a customer had a conversation with an EA about HomePrint has been considered for exclusion from this model

Results

Model #6: Overall Propensity to Participate

A stepwise logistic regression was conducted to identify **18 significantly predictive variables** for customers generally likely to enroll in any Energy Efficiency program.

	Variable Description	Source	Wald Chi Square
1	Showerheads discussed with EA	CMS	**2964.98**
2	Total number of products discussed with EA between Jan 2010 - June 2012	CMS	**2649.23**
3	HomePrint™ discussed with EA	CMS	**1848.76**
4	Account holder is a homeowner	Experian	**1424.43**
5	Account received a late payment fee, from delinquency history	CLX	**923.14**
6	Household located in region with a certain median years of education	Experian	**328.22**
7	Secondary electric service account - low voltage, typical for residential	CLX	**243.42**
8	Account is on budget billing plan	CLX	**105.09**
9	Estimated current home value	Experian	**92.50**
10	Household located in region likely comprised of individuals with blue collar occupations	Experian	**67.12**
11	Household located in region with a certain median age	Experian	**61.28**
12	Green Power discussed with EA	CMS	**54.24**
13	Average usage quantity >= 1000 kWh	CLX	**53.88**
14	Gas service account	CLX	**39.60**
15	Account located in the Tacoma weather region	CLX	**32.04**
16	Average monthly bill	CLX	**29.75**
17	Average usage quantity 150-999 kWh	CLX	**27.61**
18	Account located in region classified as "Affluent Empty Nest Households"	Prizm	**10.60**

226% more predictive than random selection

45% of participants can be found within **20%** of the population

Contents

Pilot Plan

Planned Campaigns to Test Model Efficacy

The following 4 campaigns have been confirmed for the upcoming pilot. In addition to these 4 campaigns, the Vertex team will be measuring historical results for community events. These campaigns were determined through multiple discovery sessions with program and marketing managers.

#	Campaign	Measures Marketed	Channel	Start Date	Risks	Call to Action
1	**Water Heat**	• All rebateable water heaters	• Postcard • Email TBD	• Oct 22	• Cost to customers	• Contractor referral • Call EAs • Website (dedicated webpage)
2	**Weatherization**	• Wx • Wx-Windows	• Postcard • Email TBD	• Late Oct, Early Nov	• Cost to customers	• Contractor referral • Call EAs • Website
3	**Refrigerator Decommissioning**	• Fridge Decom • Freezer Decom • Fridge/Freezer Rebates	• Postcard	• Mid-Nov	• As scheduling gets pushed, potential customer drop out	• Go to PSE.com/Recycling • Call JACO
4	**HomePrint**	• HomePrint + • All related measures and programs	• Direct mail • Outbound EAs	• Late Nov through Dec	• Brand image • Thurston county impact (EA calls) • Execution • Holidays	• Call EAs • Website

Pilot Plan

Champion/Challenger List Selection Process

Our methodology to test expected efficacy of the predictive models is detailed below:

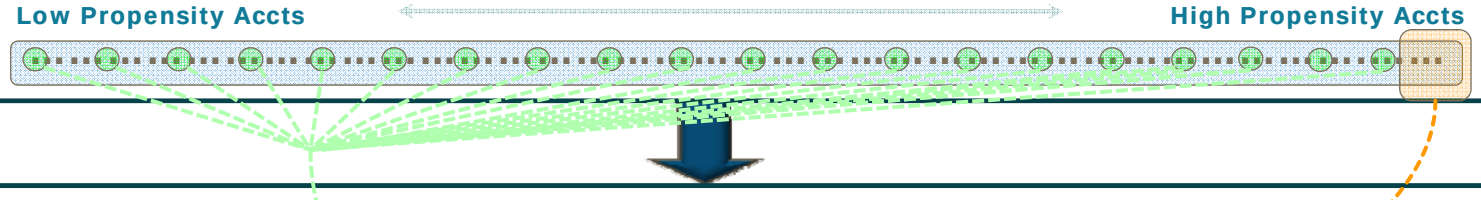
1

Identify pool of eligible applicants for campaign

- Remove ineligible account types (Water Heat program will include all eligible accounts)
- Remove accounts that have already utilized a rebate for the specific program
- Remove any other accounts or account types identified for exclusion

2

Order eligible applicant pool from lowest to highest propensity to enroll



3

Select Champion / Challenger campaign lists from pool of ordered applicants

Based on predetermined number of total campaign outgo (i.e. 60,000 postcards to be sent) select equal number of accounts for both the champion and challenger lists

- ***The Challenger list*** will be assembled by selecting the highest propensity accounts available from the pool of ordered applicants (i.e. 30,000 highest propensity accounts)
- ***The Champion list*** will be assembled by randomizing the order of the remaining eligible applicants, and executing n^{th} sampling (i.e. select every 5^{th} account from randomized, remaining applicant pool)

*** Champion List:**
Represents the process by which marketing lists are selected today

Champion List	
Account Number	Address

Challenger List	
Account Number	Address

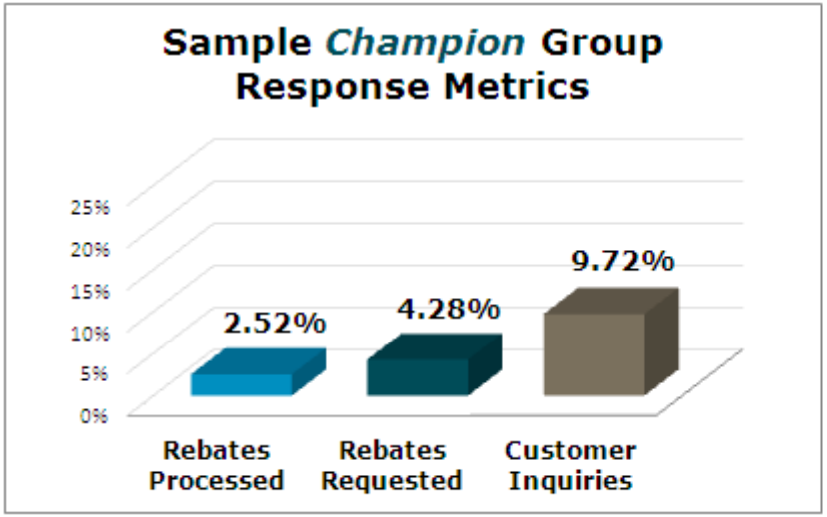
*** Challenger List:**
Represents the proposed approach for selecting marketing campaign lists going forward

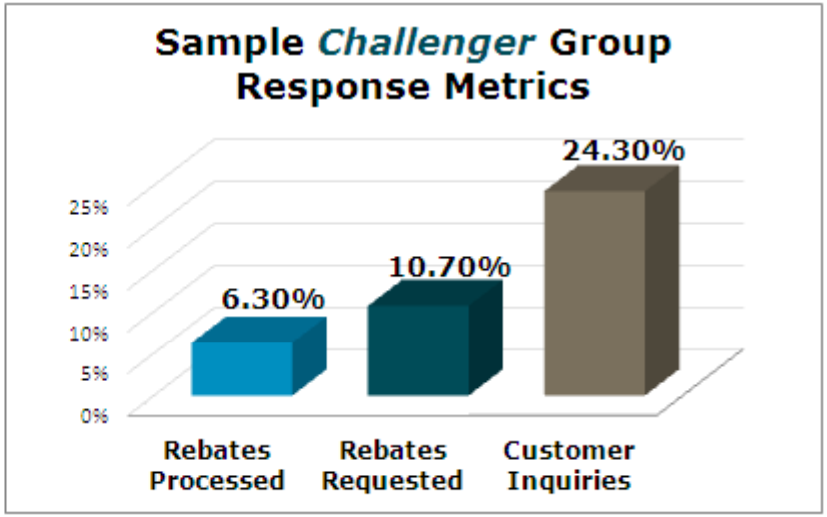
Pilot Plan

Campaign Efficacy Measurement Approach

Following campaign delivery, Vertex will begin measuring response rates for both Champion and Challenger groups. For Water Heat and Weatherization campaigns, measurement will commence approximately 90 days from the date of marketing delivery, while measurement for Refrigerator Decommissioning and HomePrint campaigns will begin 30-60 days from marketing delivery.

The following metrics will be calculated & compared between Champion & Challenger groups:





Metric	Measurement Method
Rebates Processed:	Number of rebates processed for contacted customers for marketed measures within defined campaign period / Number of customers contacted
Rebates Requested:	Number of rebates requested by contacted customers for marketed measures within defined campaign period / Number of customers contacted
Customer Inquiries:	Number of customer initiated inquiries from contacted customers regarding marketed measures within defined campaign period / Number of customers contacted

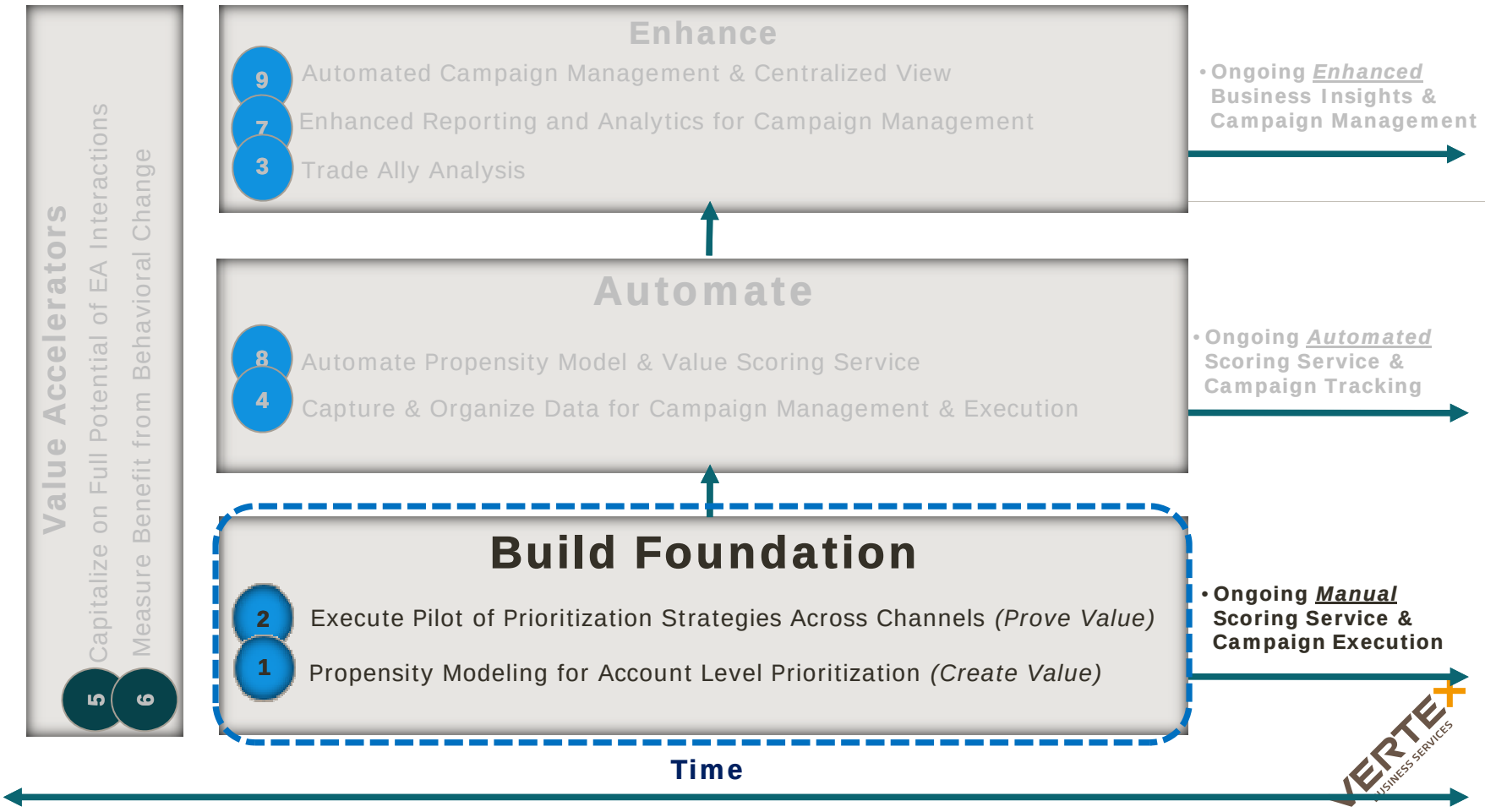
Questions?

Contents

Project Background & Status

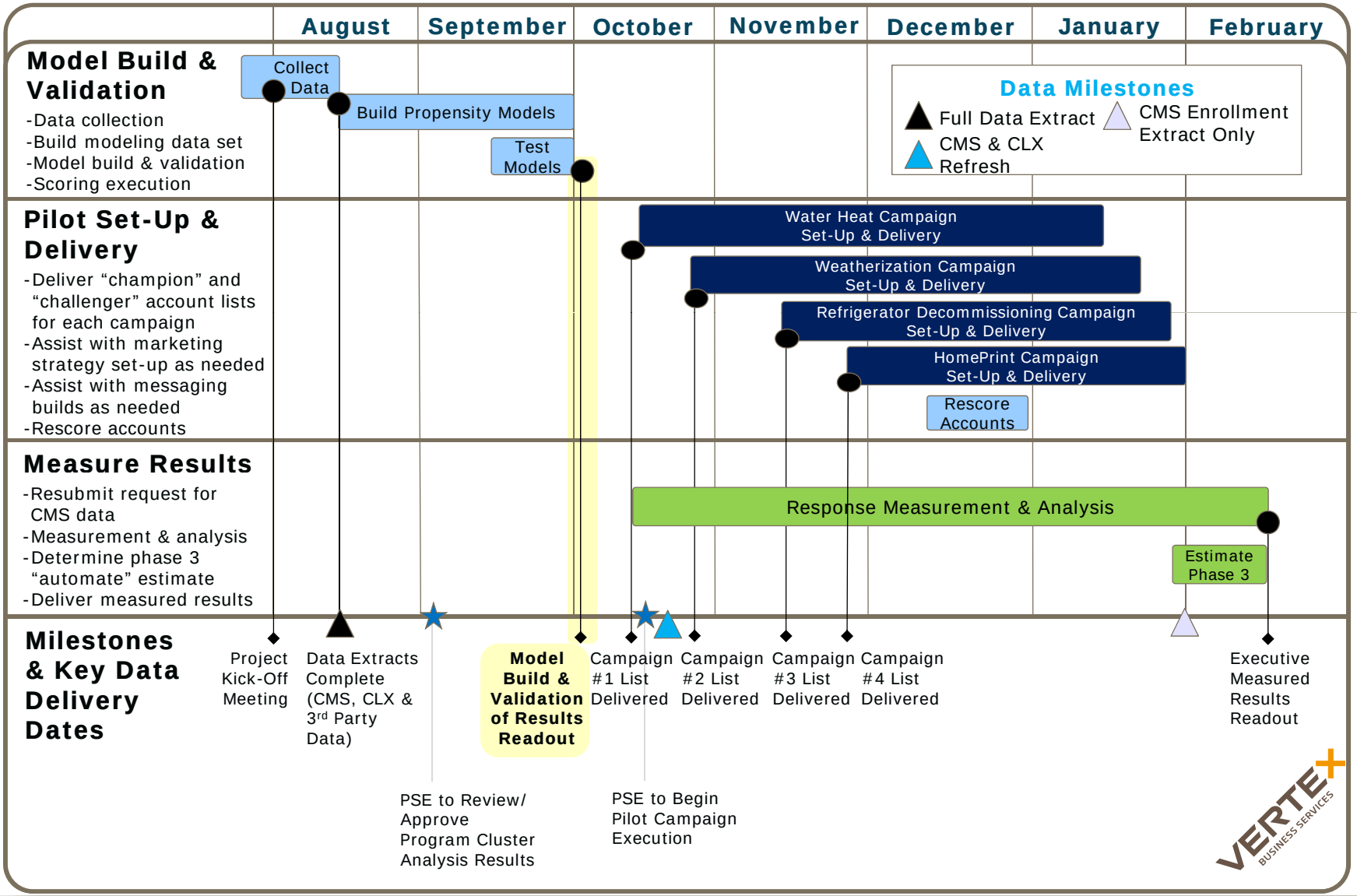
PSE Analytic Future State – Status

The objective of this effort is to build the foundation for an analytically charged Energy Efficiency program. To-date we will have created the analytical insight and value through the build and validation of predictive data models. We will prove this value in the upcoming 3 month pilot (Champion/Challenger).



Pilot Plan

Delivery Timeline



Pilot Campaign Delivery Approach

Illustrative Example – ***Why it Works…***

Current Approach Effectiveness

Champion **=** Random Account Selection

- Randomly selected accounts identified to receive marketing
- Enrollment rate typically unpredictable
- Enrollments per marketing dollar spent for comparison = .75

Before

Acct #	Propensity to Enroll in Appliance Rebates	Provided Phone Call?	Enrolled?
1	High	Yes	Yes
2	Medium	Yes	No
3	Medium	Yes	No
4	Low	Yes	No
5	High	Yes	No
6	High	Yes	Yes
7	Low	Yes	No
8	Low	Yes	No
9	Low	Yes	No
10	Medium	Yes	Yes
11	Low	No	X
12	High	No	X
13	Medium	No	X
14	High	No	X
15	Low	No	X
16	High	No	X
17	High	No	X
18	High	No	X
19	Low	No	X
20	Low	No	X

New Approach Effectiveness

Challenger = Targeted Account Selection

- Contact provided to the highest propensity to enroll customers
- Enrollments per marketing dollar spent for comparison = 1.75
- Leveraging a propensity model that is 70% effective in predicting likelihood to enroll in measure group, PSE can expect to see between **100-150% greater return on marketing dollar investment**

After

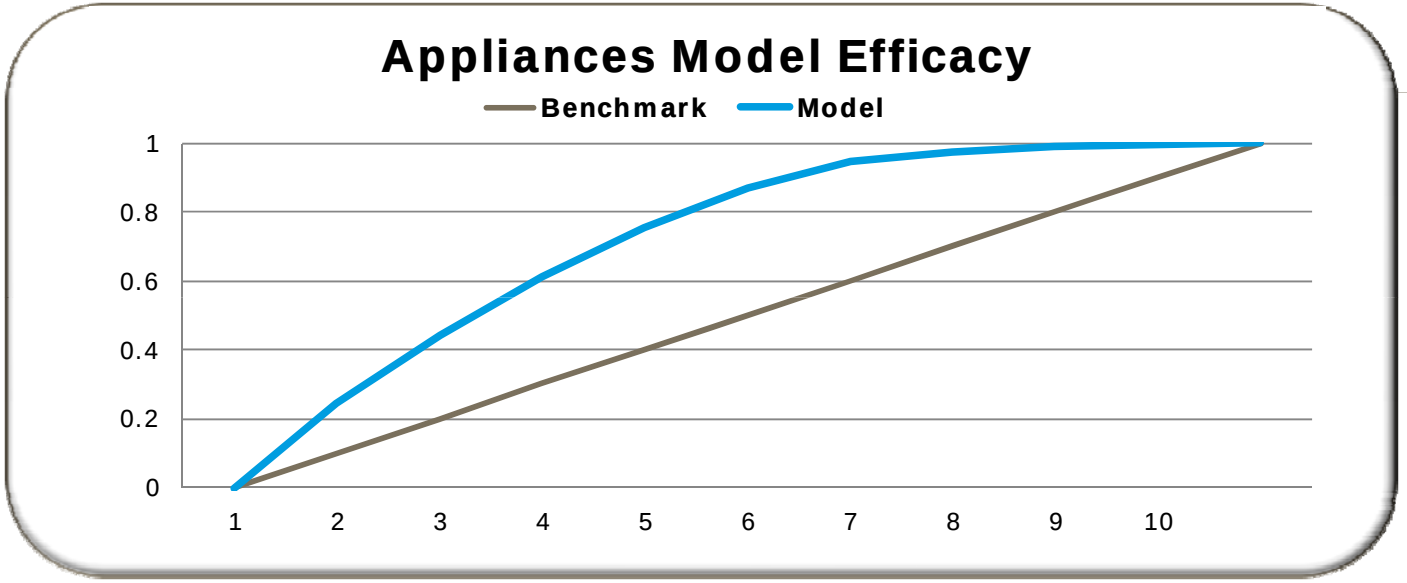
Acct #	Propensity to Enroll in Appliance Rebates	Provided Phone Call?	Enrolled?
1	High	Yes	Yes
5	High	Yes	Yes
6	High	Yes	Yes
12	High	Yes	No
14	High	Yes	Yes
16	High	Yes	Yes
17	High	Yes	Yes
18	High	Yes	No
2	Medium	Yes	Yes
3	Medium	Yes	No
10	Medium	No	X
13	Medium	No	X
4	Low	No	X
7	Low	No	X
8	Low	No	X
9	Low	No	X
11	Low	No	X
15	Low	No	X
19	Low	No	X
20	Low	No	X

Model Efficacy

Model #1: Appliances Model

Based on all of the variables selected, the propensity model is **220%** more effective in identifying customers likely to participate than the benchmark.

A random sample of 80% of cases was used to build each model using stepwise logistic regression. The remaining 20% were used as a hold out sample for testing the model. Below are the overall results.



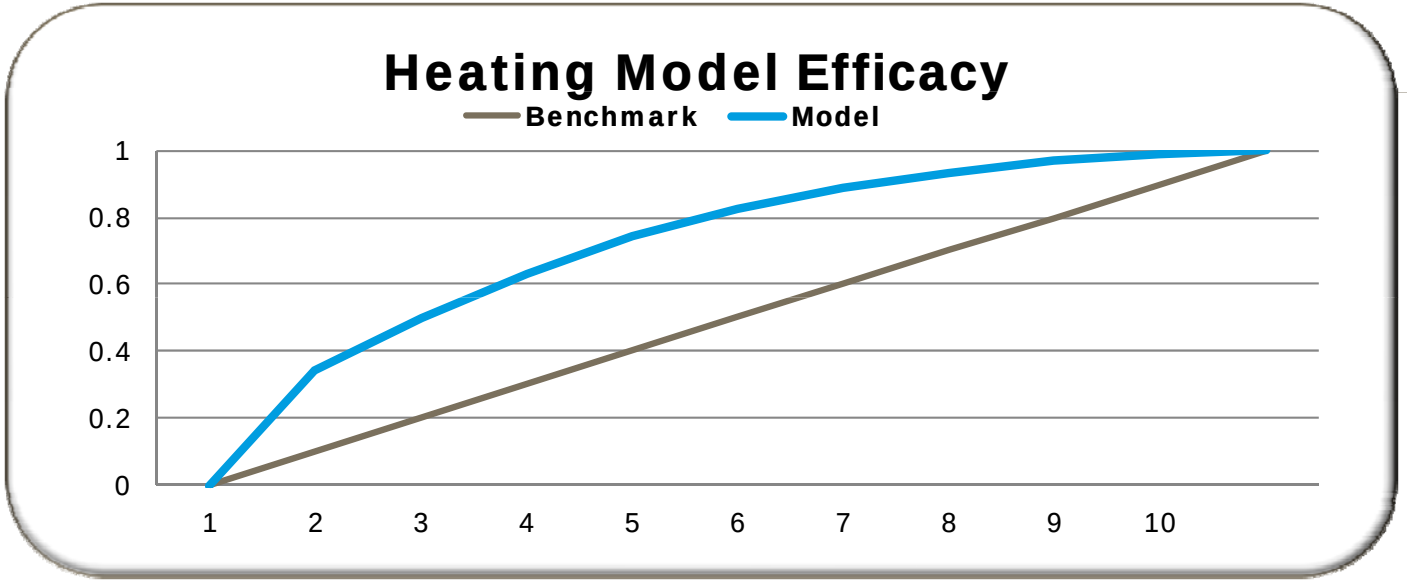
- 23 variables entered the model
- The max rescaled R square was .09
- The percentage concordant was 72.1 and discordant was 24.2
- Somer's D was .48
- Results for the hold out sample produced similar results with a Somer's D of .47

Model Efficacy

Model #2: Heating Model

Based on all of the variables selected, the propensity model is **251%** more effective in identifying customers likely to participate than the benchmark.

A random sample of 80% of cases was used to build each model using stepwise logistic regression. The remaining 20% were used as a hold out sample for testing the model. Below are the overall results.



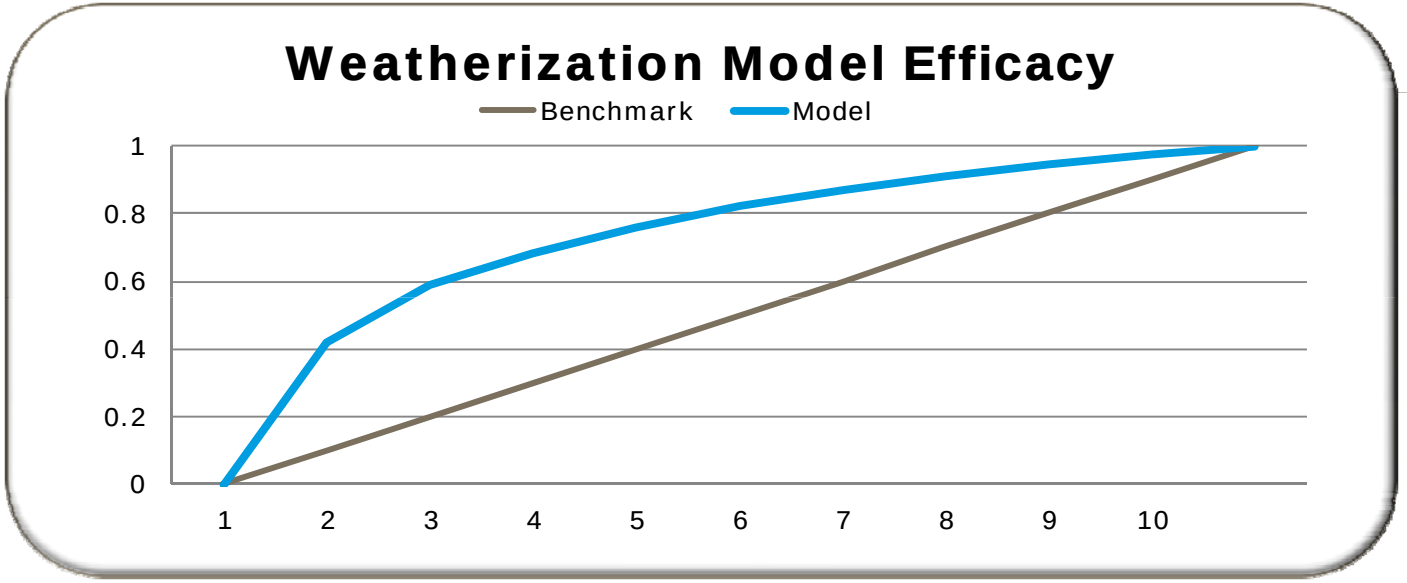
- 26 variables entered the model
- The max rescaled R square was .08
- The percentage concordant was 69.3 and discordant was 21.6
- Somer's D was .48
- Results for the hold out sample produced similar results with a Somer's D of .48

Model Efficacy

Model #3: Weatherization Model

Based on all of the variables selected, the propensity model is **293%** more effective in identifying customers likely to participate than the benchmark.

A random sample of 80% of cases was used to build each model using stepwise logistic regression. The remaining 20% were used as a hold out sample for testing the model. Below are the overall results.



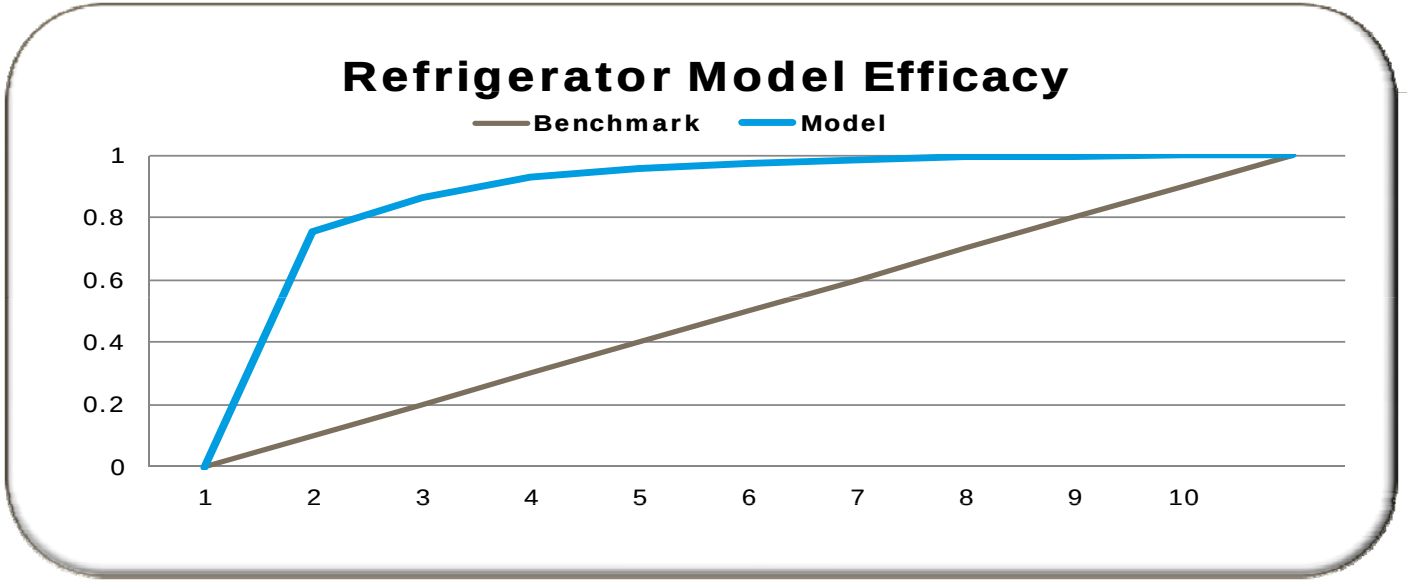
- 20 variables entered the model
- The max rescaled R square was .08
- The percentage concordant was 66.1 and discordant was 17.0
- Somer's D was .49
- Results for the hold out sample produced similar results with a Somer's D of .52

Model Efficacy

Model #4: Refrigerator Replacement Model

Based on all of the variables selected, the propensity model is **431%** more effective in identifying customers likely to participate than the benchmark.

A random sample of 80% of cases was used to build each model using stepwise logistic regression. The remaining 20% were used as a hold out sample for testing the model. Below are the overall results.



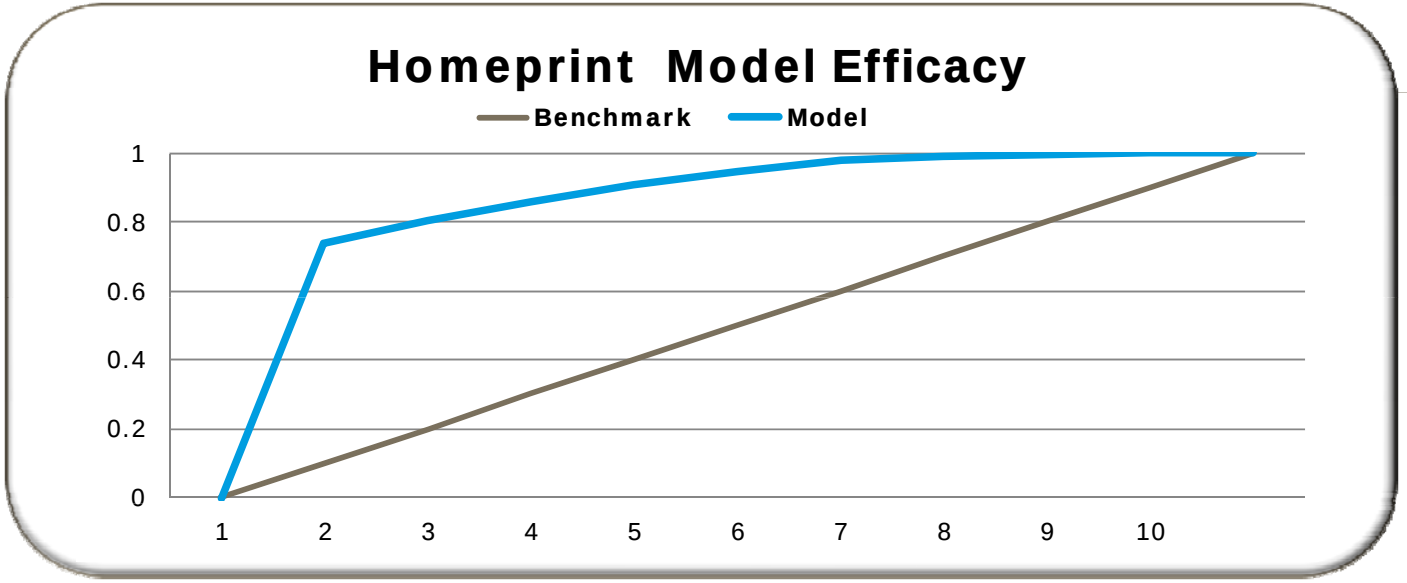
- 25 variables entered the model
- The max rescaled R square was .18
- The percentage concordant was 41.9 and discordant was 1.4
- Somer's D was .41
- Results for the hold out sample was a Somer's D of .80

Model Efficacy

Model #5: HomePrint™ Model

Based on all of the variables selected, the propensity model is **402%** more effective in identifying customers likely to participate than the benchmark.

A random sample of 80% of cases was used to build each model using stepwise logistic regression. The remaining 20% were used as a hold out sample for testing the model. Below are the overall results.



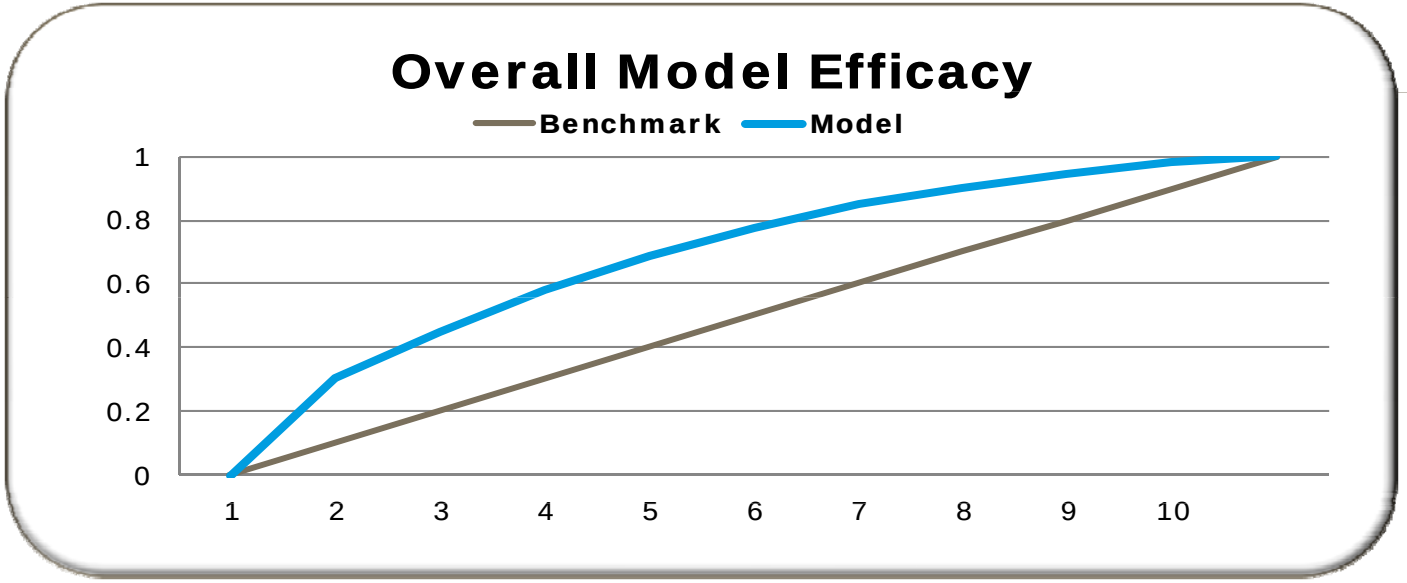
- 17 variables entered the model
- The max rescaled R square was .35
- The percentage concordant was 85.4 and discordant was 6.5
- Somer's D was .79
- Results for the hold out sample produced similar results with a Somer's D of .78

Model Efficacy

Model #6: Overall Propensity to Participate

Based on all of the variables selected, the propensity model is **226%** more effective in identifying customers likely to participate than the benchmark.

A random sample of 80% of cases was used to build each model using stepwise logistic regression. The remaining 20% were used as a hold out sample for testing the model. Below are the overall results.



- 19 variables entered the model
- The max rescaled R square was .10
- The percentage concordant was 70.4 and discordant was 27.6
- Somer's D was .43
- Results for the hold out sample produced similar results with a Somer's D of .41

Vertex Analytic Team

Engagement Lead

Hugh Burgin

Project Management

Micah Dehenau

Statisticians

Robert Stephan

Lynn Cherry

Consultants

Kanae Wilcox-Perrotta

Angela Ray

Conclusion

For further discussion, please contact:

Micah Dehenau at 248.410.3132