Northeast

Newsletter opt-out analysis

Rashmi Belur • Caitlin Blanchard • Jazmine Estrada • Scott Hawes • Nathan Millette • Donna Mowry • Alec Schuberth • John Templeman

Contents

Background

- AAA Northeast has a strategic relationship with Bryant University to assist in the advancement of their data science capabilities
- Data science has been identified as a key driver of business success
- We were given two files to begin the process of assisting AAA Northeast to leverage data for profiling and predictive analysis
  - Monthly print newsletter opt outs
  - Sprint members

Objectives
Group 1

Analyze members that have opted to receive digital communications rather than mail

- Provide a profile of the members that have opted out in order to better understand and communicate to them
- Develop predictive (look-alike) models to assist targeted marketing efforts to increase the number of members that receive electronic communication

Approach

1. Obtained data set and performed initial analysis
2. Met with client to review business and discuss early identified data patterns
3. Transformed data set to prepare for analytics methods
4. Prepared customer profile for each objective
5. Created two predictive models for the objective:
   - Decision Tree
   - Logistic Regression

Analysis Results:

Opt Out – Demographic Profile

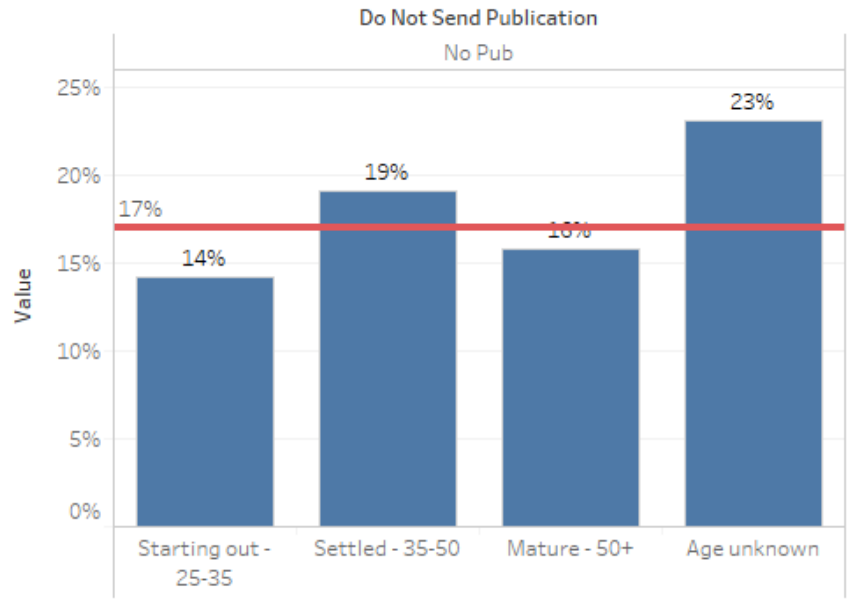
Strong identifiers :

- Age = Unknown
- Gender = Unknown
- Residence = Multi Family

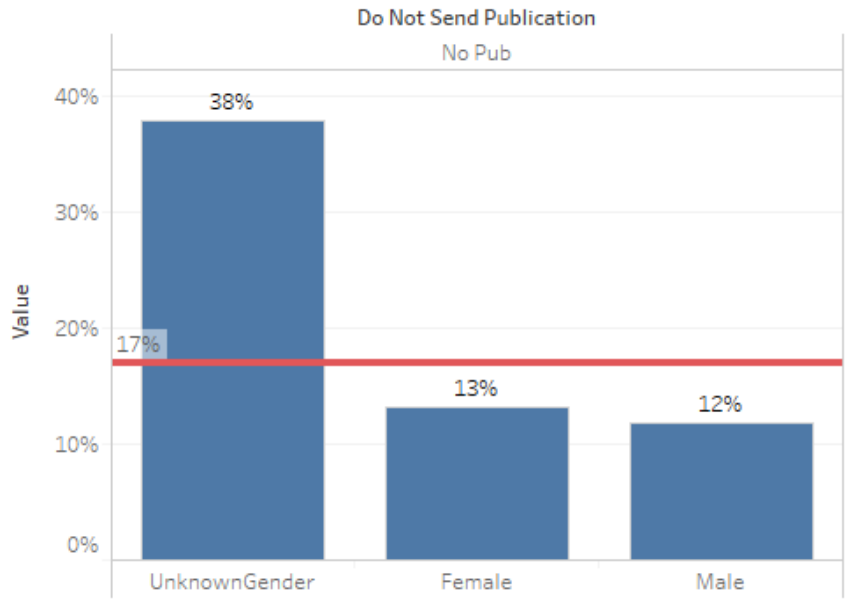
No Stand Out Features for the following dimensions:

- Language
- Number of Children
- Income
- Education
- Home Owner Status

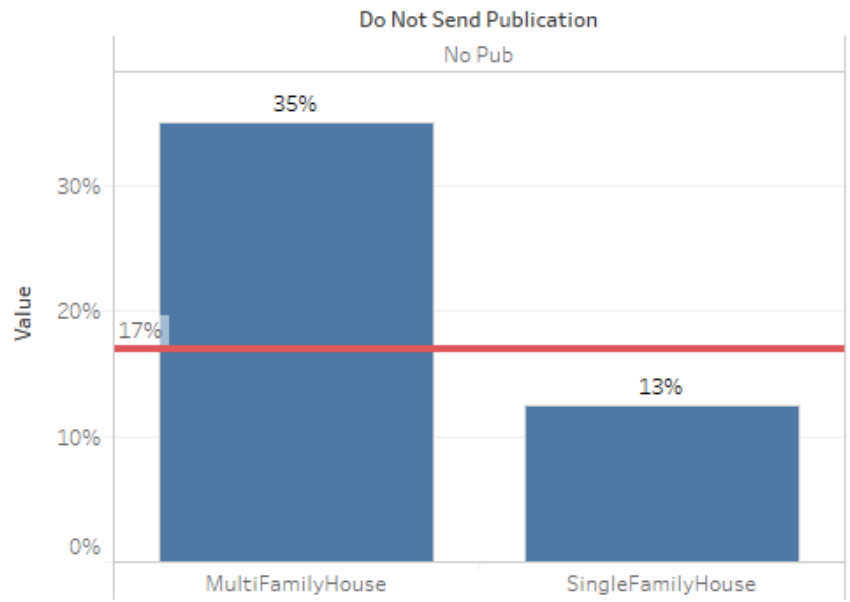
Member Age



Gender



DwellingType



AAA Northeast

Analysis Results:

Opt Out – Customer Preferences Profile

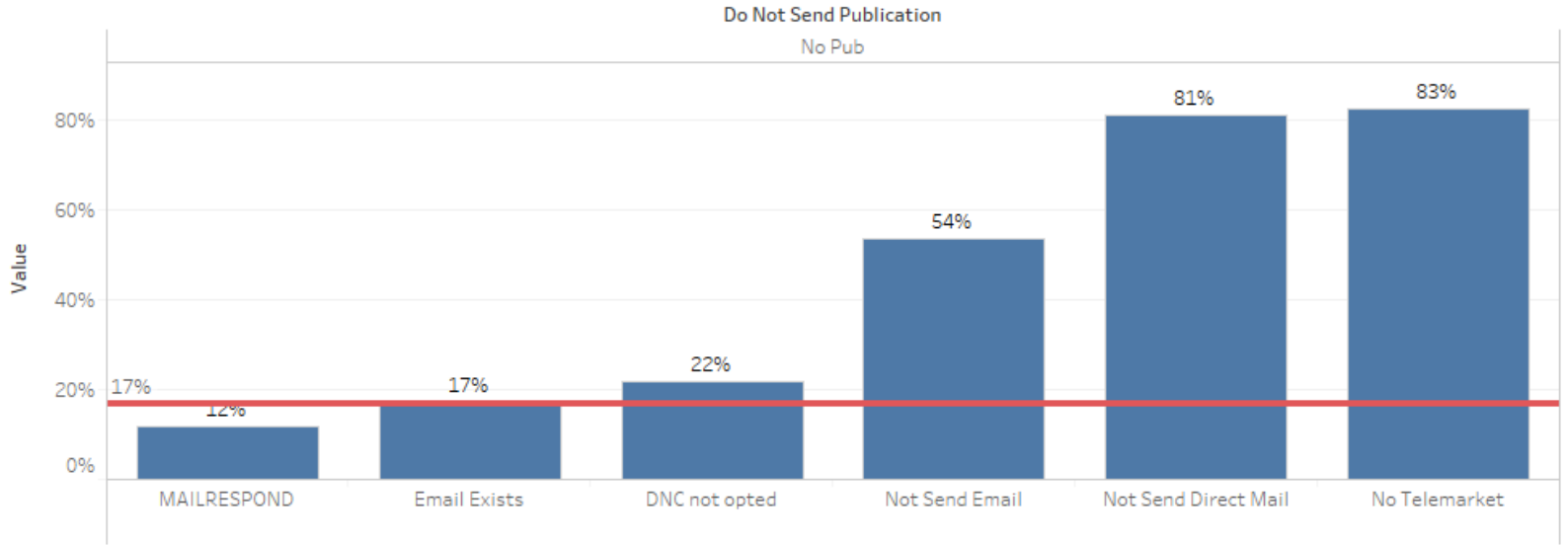
Strong identifiers:

- Do Not Email
- Do Not Send Direct Mail
- Do Not Tele-market
- Moderate correlation to Plus and Premier Plan Types

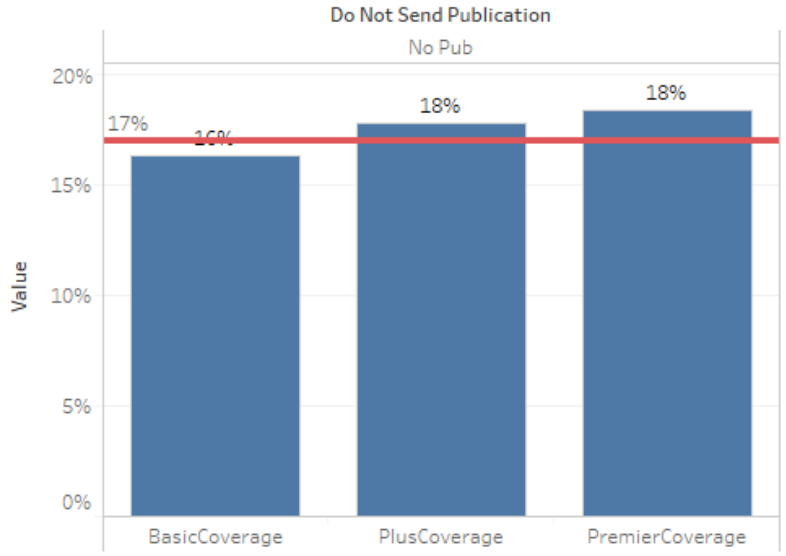
No Stand Out Features for the following dimension:

- Renewal Method

Preferences



PlanType



AAA Northeast

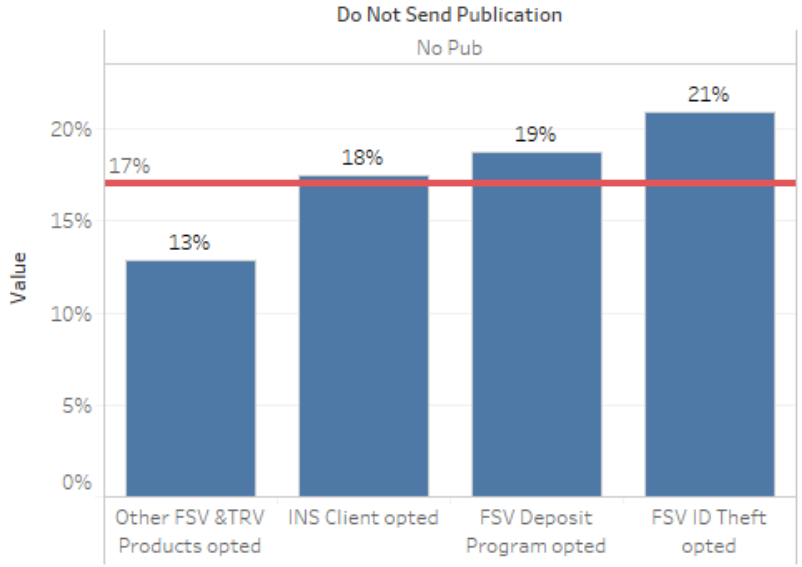
Analysis Results:

Opt Out – Customer Behavior Profile

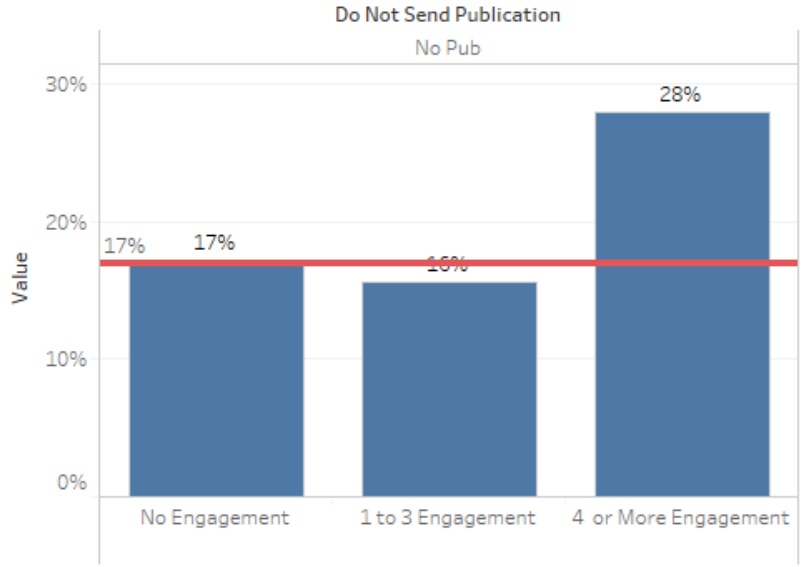
Strong identifiers:

- Enrolled Products = FSV
- Engagement in Number of Services >= 3 Products
- Membership Length => 10 years
- Last Transaction = Billing

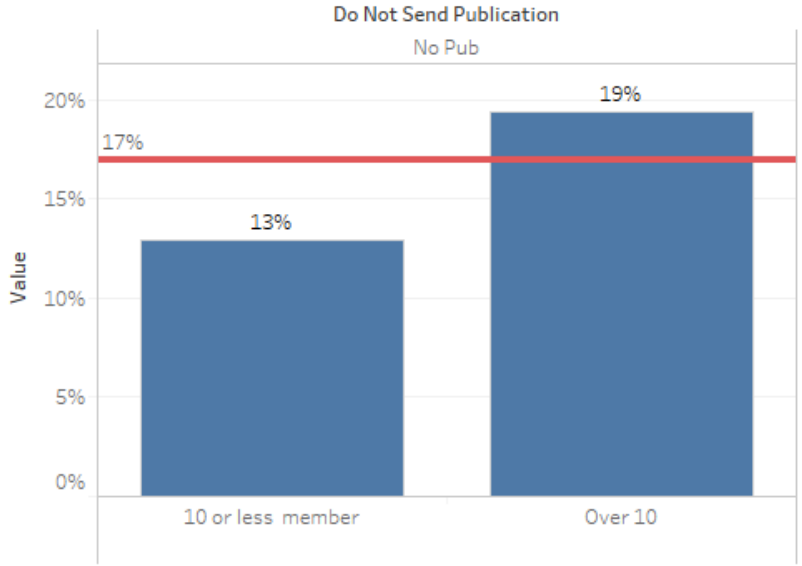
Enrolled Products



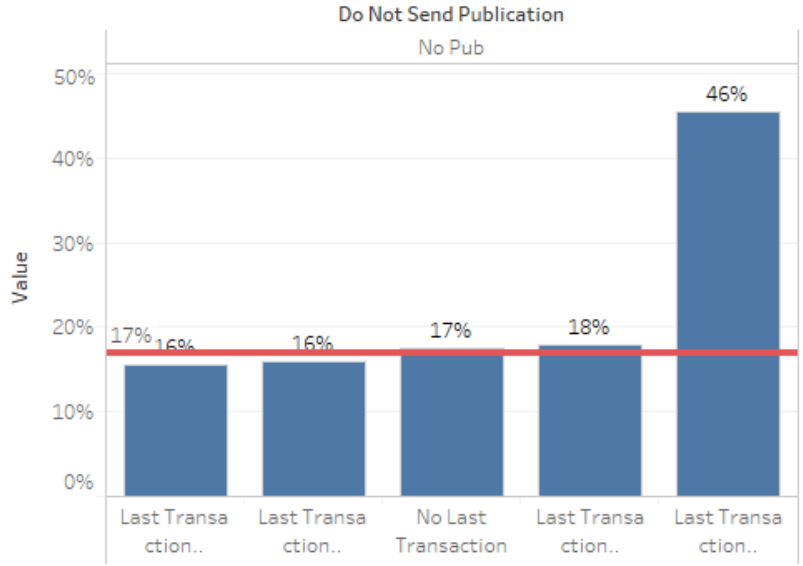
Engagement



Membership Age



Last Transaction



Decision Tree:

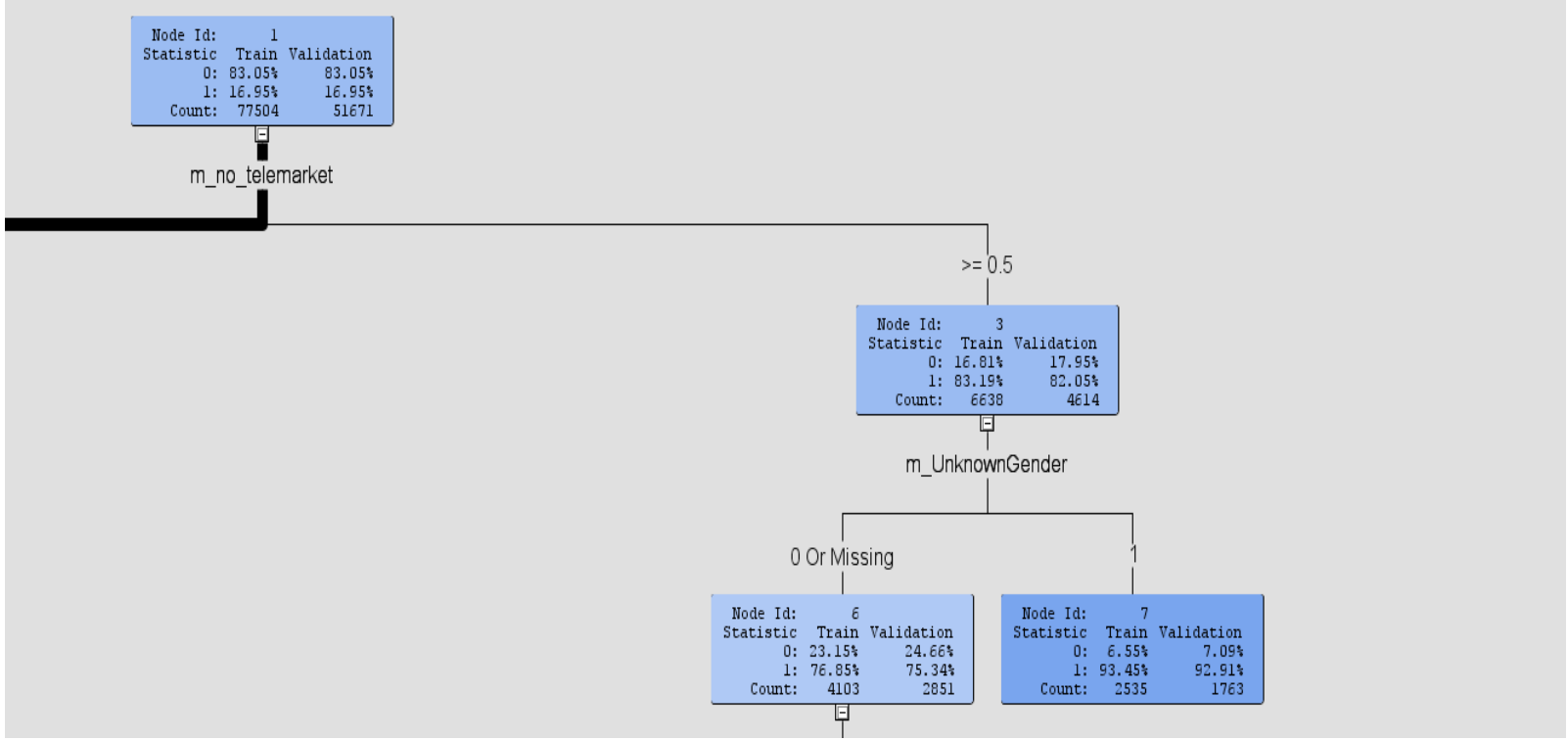
Opt Out - Condensed

Critical variables in order of importance

1. Do Not Telemarket
2. Unknown Gender

Other important variables

1. Multi-Family Dwelling
2. Length of Membership > 9 years
3. Do Not Send Email
4. Age < 50 years OR unknown
5. Last Transaction is Mail



Logistics Regression :

Variable Coefficients by Importance

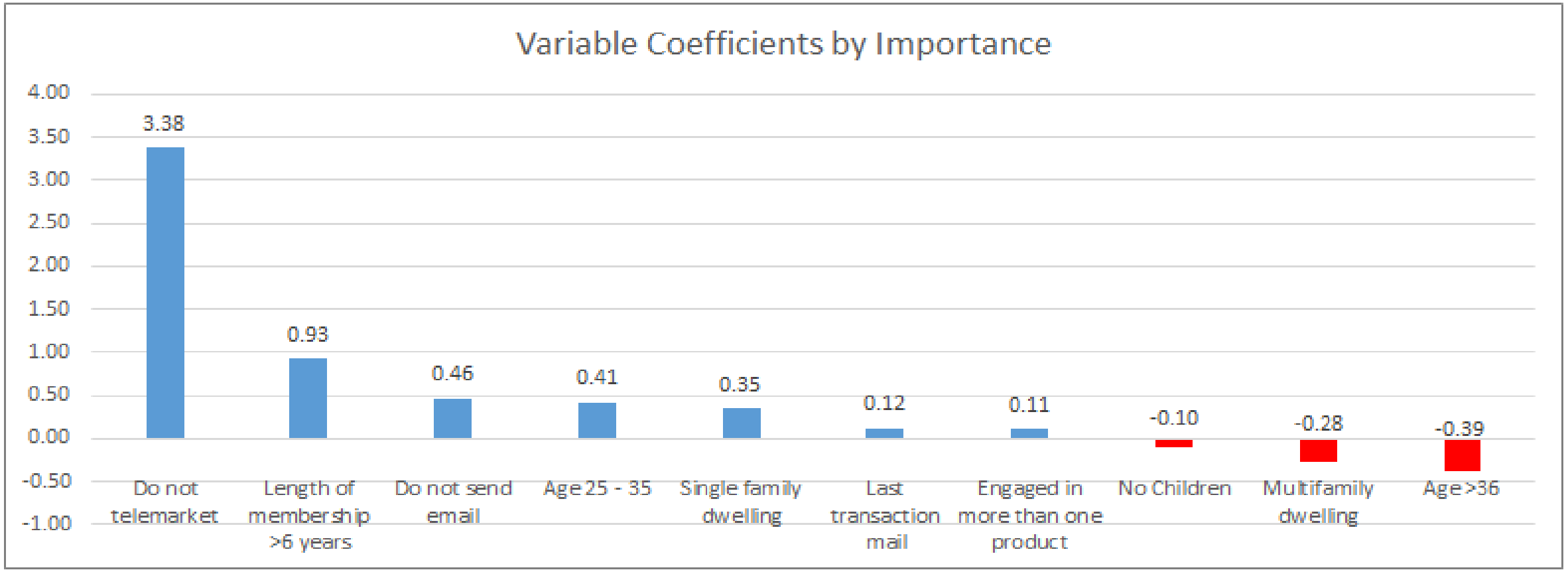
No Children Age 25 - 35 Single family dwelling

Last transaction mail Engaged in more than one product

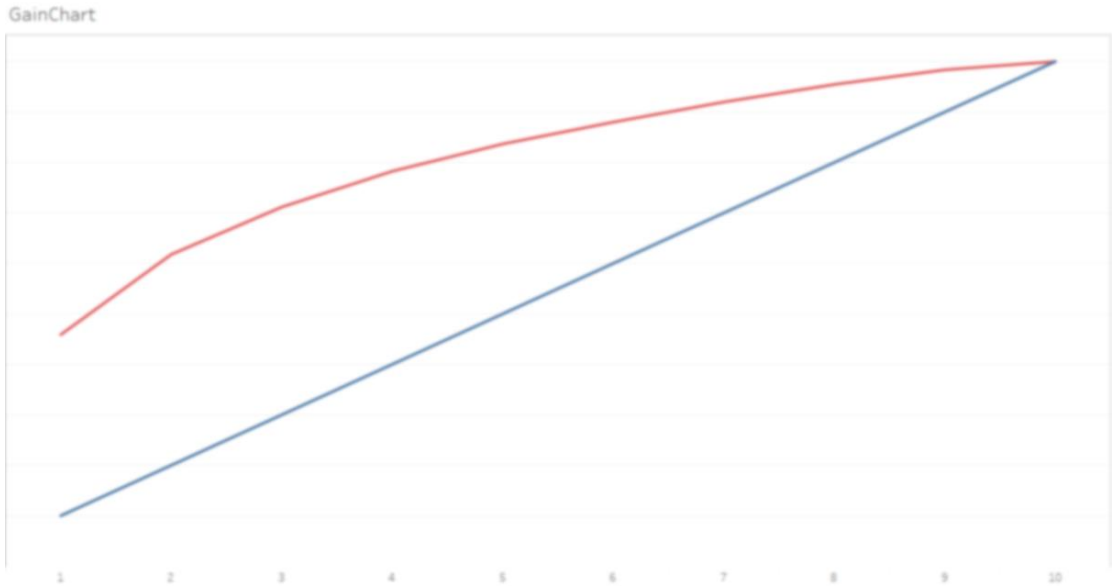
Do not telemarket

Do not send email Length of membership >6 years

Age >36 Multifamily dwelling



Decile Analysis



1st Decile has 78.6% response rate compare to overall 17% with increase of 363% from random selection

Decile Analysis

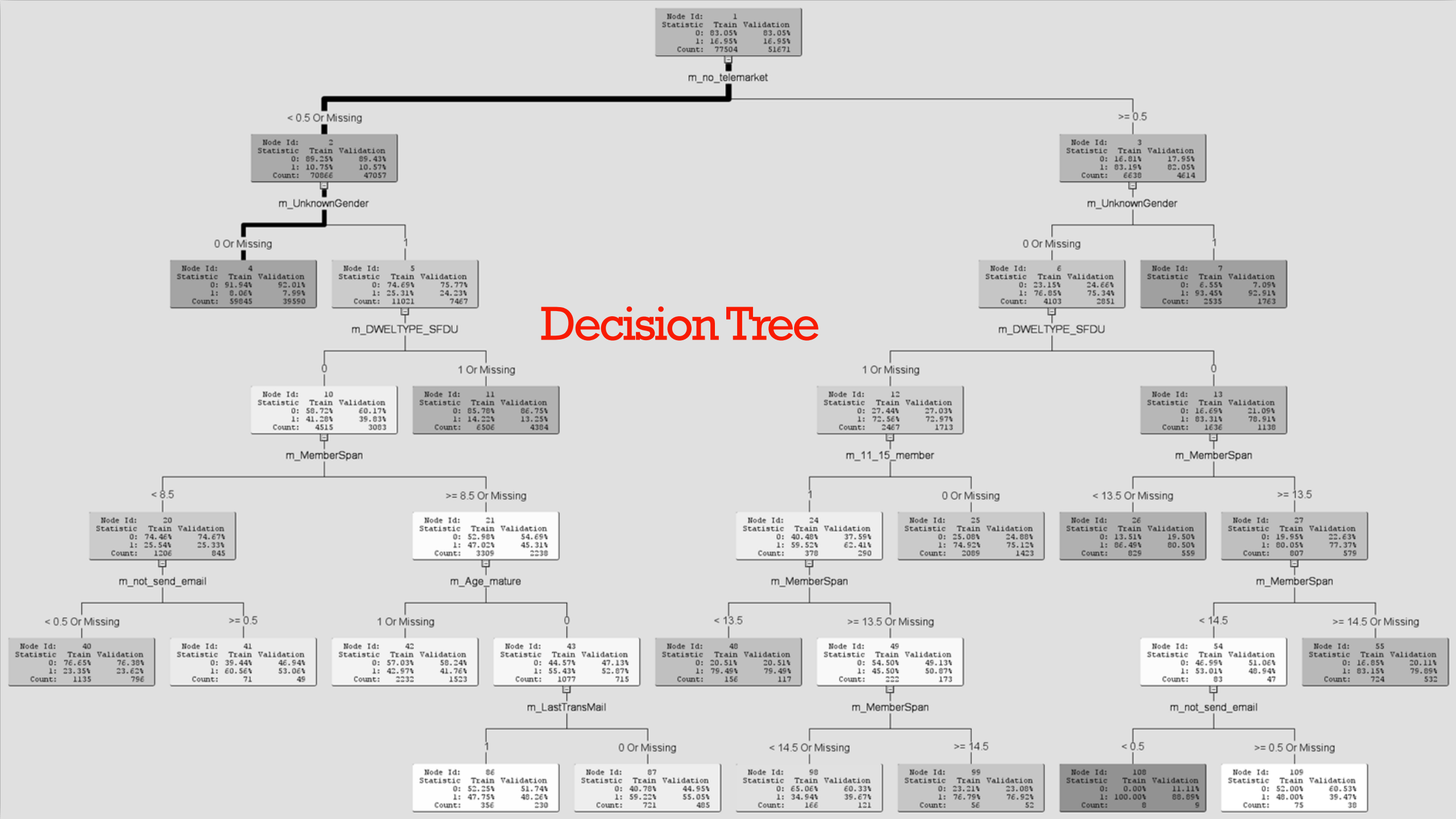
Decile	Members	Do Not Send Publication	%Do Not Send Publication	Cumulative % Do Not Send Publicati..	Lift	Cumulative Lift
1	12,831	10,090	78.6%	78.6%	4.63	4.63
2	12,832	3,424	26.7%	52.7%	1.57	3.10
3	12,832	1,720	13.4%	39.6%	0.79	2.33
4	12,832	1,454	11.3%	32.5%	0.67	1.91
5-9	64,160	4,873	7.6%	18.7%	0.45	1.10
10	12,832	241	1.9%	17.0%	0.11	1.00
Grand Total	128,319	21,802	17.0%	17.0%	1.00	1.00

Recommendations and Next Steps:

1. Research unknown data like Age, Gender and Dwelling Type
2. Consider providing incentives to customers during data collection
3. Run full AAA Data Set through the Logistic Regression Predictive Model to obtain list of best candidates to Opt Out of Newsletter
4. Provide instruction to members on how to receive electronic newsletter

Appendix

- Decision Tree
- Logistic Regression equation
- Analysis of Maximum Likelihood Estimates



Decision Tree

Logistic Regression Equation

$$P_{(\text{opt out})} = \frac{e^{-2.36 + 3.38x_1 + .93x_2 + .46x_3 + .41x_4 + .35x_5 + .12x_6 + .11x_7 - .1x_8 - .28x_9 - .39x_{10}}}{1 + e^{-2.36 + 3.38x_1 + .93x_2 + .46x_3 + .41x_4 + .35x_5 + .12x_6 + .11x_7 - .1x_8 - .28x_9 - .39x_{10}}}$$

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept		1	-2.3636	0.0899	690.95	<.0001		0.094
m_1plus_engage	0	1	0.1082	0.0160	45.52	<.0001		1.114
m_6_over_member		1	0.9289	0.0348	714.19	<.0001	0.2140	2.532
m_Age_Settled_mature		1	-0.3816	0.0381	100.53	<.0001	-0.0677	0.683
m_Age_Starting	0	1	0.4132	0.0433	90.87	<.0001		1.512
m_DWELTYPE_MultiFamily	0	1	-0.2752	0.0740	13.84	0.0002		0.759
m_DWELTYPE_SFDU	0	1	0.3501	0.0121	832.36	<.0001		1.419
m_LastTransMail	0	1	0.1214	0.0125	94.16	<.0001		1.129
m_NumChild_0	0	1	-0.1055	0.0145	52.82	<.0001		0.900
m_UnknownGender	0	1	-0.7028	0.0156	2037.25	<.0001		0.495
m_no_telemarket		1	3.3818	0.0482	4930.54	<.0001	0.5256	29.423
m_not_send_email		1	0.4597	0.0397	134.19	<.0001	0.0871	1.584