To: The Client
From: 662761431
Date: 11/24/25
Re: Motley Fool Newsletter Signup Model

This memo is in response to your request to build a model for identifying prospects most likely to sign up for the Motley Fool investment newsletter.

Specifically, you requested a model and analysis that would address four questions:

1. Can we build a model to cut our mailing quantities by 25% and still get most of our responses?
2. What are the variables used in the model?
3. Which of those variables are the most impactful and what is their relationship with the probability of making a sale?
4. What is an alternative cut-point for the prospect universe?

Results show that the top 75% of prospects contain 83% of the total sign-ups. The variables used in the model are listed in Table 1. The most impactful variables were the number of private third-party insurance policies, being in the Large Family, Employed Child segment, and the average age in the neighborhood. The first two had a positive association with purchase, while average age in the neighborhood showed a negative association. Smaller but meaningful predictors included Social Class C, Pct. Married in the Neighborhood, Household with 2 Cars, and Social Class B1. An alternative cut-point exists at 54% of the prospect universe which contains 71% of the sign-ups. This cut-point was identified using lift analysis as discussed below in page 2.

Top 75% of Prospects. Regression modeling was used to identify the characteristics that were most strongly associated with signing up for the Motley Fool newsletter. The results of that modeling showed that the top 75% of prospects contained 83% of the total sign-ups. By abstaining from contacting the bottom quarter of the prospect file you can increase your response rate from 5.7% to 6.3%.

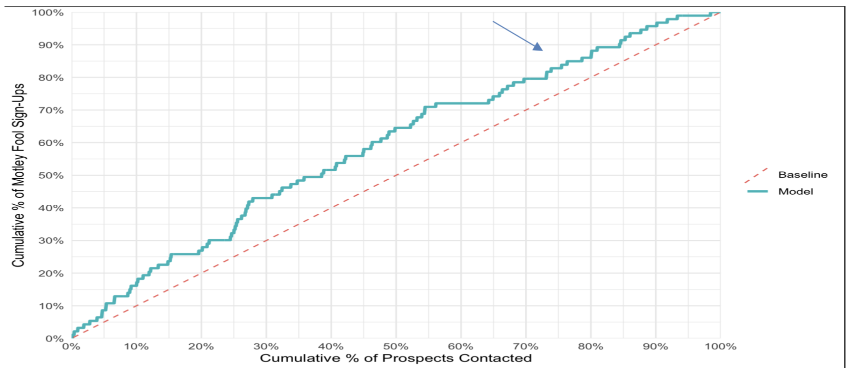
Variables in the Model and Impact. The variables used in the model for predicting sign-ups for the Motley Fool investment newsletter are shown in Table 1. The slope for Number of Private Third–Party Insurance Policies is 17.795%. This means that for each additional private third-party insurance policy a prospect holds, the predicted probability that they sign up for the Motley Fool newsletter increases by an estimated 17.795 percentage points, holding all other variables constant. The next most impactful variable is membership in the Large Family, Employed Child segment, which increases the predicted probability of newsletter sign-up by 9.130 percentage points. Average Age in the Neighborhood has a negative relationship with sign-up; each unit increase decreases the predicted probability of signing up by 1.258 percentage points.

Table 1

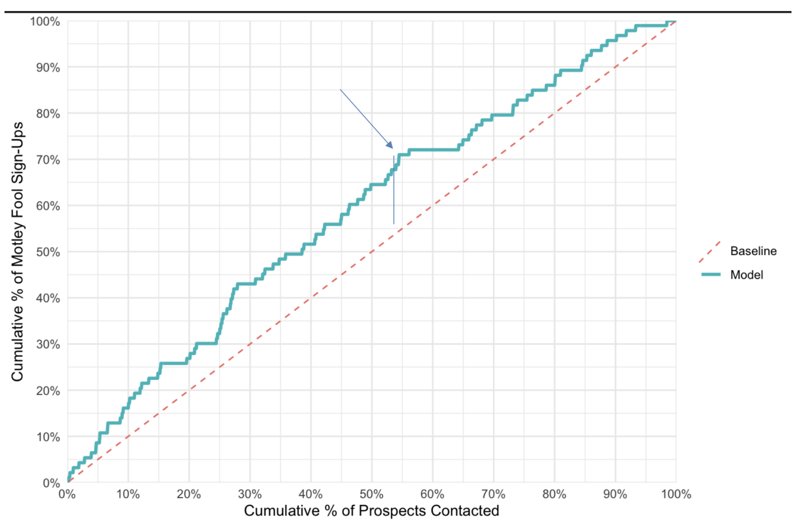
Variable	Impact on Probability of Sale
Number of Private Third–Party Insurance Policies	17.795%
Large Family, Employed Child	9.130%
Average Age in Neighborhood	-1.258%
Social Class C	-0.134%
Pct. Married in Neighborhood	0.084%
Household with 2 Cars	0.079%
Social Class B1	0.062%

Gains Chart. The Gains Chart from the model is shown below. The orange line labeled 'Baseline' shows the percent of Motley Fool newsletter signups if prospects were selected on a random basis. That is, we would expect that a random selection of 10% of all prospects to contain 10% of sign-ups; 20% of randomly selected prospects would account for 20% of sign-ups, etc.

The line labeled "Model" shows the expected results if prospects were contacted in order of their predicted probability of signing up for the Motley Fool Investment Newsletter, as estimated by our general linear model. You can see that the top 10% of prospects account for around 17% of all sign-ups. The spot on the gains chart marked with an arrow shows that contacting the top 75% of prospects captures 83% of all Motley Fool newsletter sign-ups.



Alternative Cut-Point. Lift is defined as the difference in cumulative percent of responses from the model vs. the baseline of using random selection. A sensible place to cut the number of contacts is where the lift is maximized since that point gives you maximum benefit from using the model. The gains chart is repeated below with the maximum lift marked. This occurs at 54% of prospects and 71% of sign-ups.



Summary. You can reduce the number of contacts by 25% while still capturing 83% of all Motley Fool newsletter sign-ups. The most impactful variables in predicting sign-up were the Number of Private Third-Party Insurance Policies, membership in the Large Family, Employed Child segment, Average Age in the Neighborhood, and Social Class C. The first two variables had a strong positive association with sign-up, while the latter two showed a negative association. An alternative cut-point based on maximum lift occurs at 54% of prospects and contains 71% of the total sign-ups.

Technical Appendix

This technical appendix provides details as to how the data was prepared for modeling and the construction of the model itself.

Logistic Regression. A logistic regression model was built to identify the characteristics of prospects more likely to sign up for the Motley Fool investment newsletter. The results of that model are shown below. Details regarding the steps preceding the actual model construction follow.

Table 2

	Estimate	Std..Error	z.value	Pr...z..
(Intercept)	-3.235999042	0.324514481	-9.971817	2.024898e-23
AWAPAR	1.404374004	0.336438193	4.174241	2.989815e-05
MSKC	-0.019121889	0.008274030	-2.311073	2.082882e-02
MOSTYP_34	0.892806064	0.356685637	2.503061	1.231242e-02
MRELGE	0.016539707	0.007743598	2.135920	3.268592e-02
MGEMLE	-0.311974496	0.138894880	-2.246119	2.469636e-02
MAUT2	0.008786891	0.003978404	2.208647	2.719921e-02
MSKB1	0.006964229	0.003461996	2.011622	4.425976e-02

The p-values for each of the variables are shown in the rightmost column. As you can see, all the p-values fall below the 5% threshold for statistical significance. This means that, at the 5% level, each variable included in the model has a statistically significant association with the likelihood of signing up for the Motley Fool newsletter.

Data Preparation and Variable Selection. The initial file contained 5,403 rows and 28 variables. A number of these variables required adjustment prior to building the model. A breakdown of said variables is provided in the next page.

Ordinal Variables. The file contained 13 geo-demographic ordinal variables using the L3 coding system. That is, a particular value for one of these variables represented a range of percentages of people of a certain type in the prospect's neighborhood. To use these variables in a linear model, the original values were re-scaled to the midpoint of the percentage range. Similarly, the file contained 5 spending variables using the L4 coding system, where each value represented a range of amounts a prospect spent on a product category. These were also re-scaled to the midpoint of the spending range, as shown in Table 3.

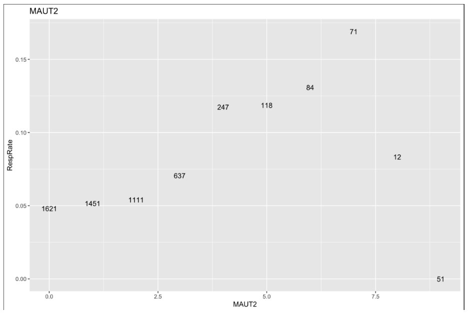
Table 3

L3: Geo Demographic Format			**L4: Spend Format**		
Variable Value	**Original Value**	**New Value**	**Variable Value**	**Original Value**	**New Value**
0	0%	0	0	\$0	\$0
1	1–10%	5.5%	1	\$1–\$49	\$25
2	11–23%	17%	2	\$50–\$99	\$75
3	24–36%	27%	3	\$100–\$199	\$150
4	37–49%	40%	4	\$200–\$499	\$350
5	50–62%	52%	5	\$500–\$999	\$750
6	63–75%	65%	6	\$1,000–\$4,999	\$3,000
7	76–88%	77%	7	\$5,000–\$9,999	\$7,500
8	89–99%	90%	8	\$10,000–\$19,999	\$15,000
9	100%	100%	9	\$20,000+	\$30,000

Categorical Variables. Two categorical variables, MOSTYP and MOSHOO, were included in the original file. Because these variables contain coded categories, each category was converted into its own 0/1 indicator variable so it could be used properly in the model. For example, if an entry had the MOSHOO variable be equal to 2, then that would translate to the MOSHOO2 variable being equal to 1 for that specific entry, and all other MOSHOO indicators being 0.

Holdout Sample. We set aside 30% of the records at the start and did not use these in building the model. This holdout group acts like "new customers" and lets us see how well the model would work if applied to a larger prospect universe. Testing our model on data outside of the training set allows us to check for overfitting, which is when the model matches so closely to the data that it practically memorizes it and fails to make predictions for new data. By checking for overfitting, we are making sure that the model's performance stays consistent even when introduced to more data points.

Non-Linear Relationships. Graphs showing the response rate for each of the quantitative variables were created. For those variables that appeared to have a non-linear relationship with response, we examined both linear and quadratic forms. For example, the graphs below show the pattern of response rate for MAUT2 and MFALLE. Individual logistic regression models for a linear vs. quadratic relationship with response were calculated, and the AIC values were examined, summarized in Table 4.



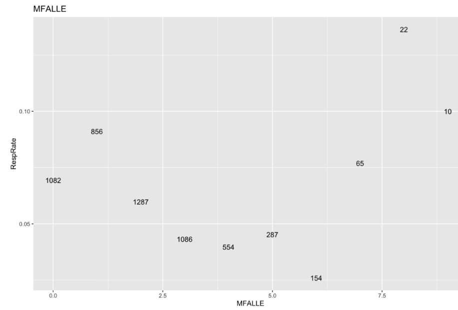


Table 4

Independent Variable	**Linear AIC**	**Quadratic AIC**
MAUT2	2,439	2,448
MFALLE	2,449	2,455

The AIC for the linear model was 2,439 vs. 2,448 for MAUT2, and therefore the linear form was used as a candidate independent variable for logistic regression. Similarly, the AIC for the linear model for MFALLE was 2,449 vs. 2,455, and therefore the linear form was used as well.

Logistic Regression Model. After preparing all the variables for potential use in the model, all 26 variables were submitted to a logistic regression model using stepwise variable selection. This resulted in 7 statistically significant variables as shown in Table 2.

Linear Regression Model. After using logistic regression to select the variables for the model, the final set of independent variables was used to build a linear regression equation. This was done to obtain the slope coefficients for each variable, which quantify the change in the probability of signing up for the Motley Fool investment newsletter associated with a one-unit change in that variable. The coefficients from that model were used to compute the impact percentages that appear in Table 1.

Table 1

Variable	**Impact on Probability of Sale**
Number of Private Third–Party Insurance Policies	17.795%
Large Family, Employed Child	9.130%
Average Age in Neighborhood	-1.258%
Social Class C	-0.134%
Pct. Married in Neighborhood	0.084%
Household with 2 Cars	0.079%
Social Class B1	0.062%

Elan Shlain.