An Analysis of Market Response to the USB Toaster

Skip Potts, Abby Shambaugh, Jordan Majoros, Risheng Li

3/21/2021

Abstract

The purpose of this report is to use confidence intervals, logistic regression, as well as analysis of contingency tables in order to find the likelihood that the USB Toaster will be a successful product, useful demographic points to aid in formulating a successful marketing campaign, and potential profit or loss. By calculating credible intervals of the Bayesian mean for the proposed five major markets, we found that at least 95% of the time the probability of profit was essentially 0% for this market. Subsequent analysis using logistic regression and contingency tables identified a potentially profitable and larger market of college students for the USB Toaster in which the average expected market share was expected to fall between 26.2% and 32.2% using a 95% confidence interval. Given this interval of expected mean market share, the USB Toaster has a strong chance of being profitable in the college market contingent on the fixed sale price and the feasibility of the recommended changes concerning marketing and distribution.

Introduction

The client has requested that we evaluate the profit potential of the USB Toaster for the five major metropolitan areas of Toronto, ON; New York, NY; Philadelphia, PA; Dallas, TX; and San Francisco, CA. Unfortunately, with the required break even point of 24.13% sales, even a cursory look at the data informs an analyst that this is not a profitable proposition. Thus, the initial focus of analysis was shifted from using the data at hand to determine the profitability of the USB Toaster in the given market, to determining if a market existed where the product could feasibly be profitable.

Here, we have rendered a detailed analysis of the proposed new product, the USB Toaster, based on the comprehensive data gathered by our firm and the summary data given to us by the client. We have compiled an examination of the data collected in both samples and highlighted which demographic data points may be of interest, and most importantly, we predicted how successful the USB Toaster potentially might be in the identified market based on our comprehensive model. The market demographic that was identified as having the highest profit potential was college students. While it is understood that a shift to this highly focused market represents a significant change in strategy for the client, we believe this is not only the single market that the USB Toaster may find success in, but also provides a target demographic to pursue that clearly can lead to a profitable product roll-out dependent on other contingent factors (discussed in more depth later).

The processes of reaching the above objectives will be detailed in sections covering the data, exploratory data analysis, the model, and discussion in this report. During the course of the paper, we will also review the results of our analyses into the original Specific Aims noted below.

Specific Aim 1 (SA1): How are the two datasets (from Table 1 and Table 2) different and why?
Specific Aim 2 (SA2): What is the likelihood, based on all available data, that the USB Toaster will be a success?'
Specific Aim 3 (SA3): What are some useful demographic data points that might help inform a potential marketing campaign?
Specific Aim 4 (SA4): What does the data tell us in terms of potential profit or loss based on the outcome of the analysis?

The Data

The Two Data Sources

The data we have analyzed comes from two different sources, the client and the consulting firm. The data obtained by the client provides valuable information on demographics that may or may not have interest in purchasing the USB Toaster, although it is clearly not representative of the population. More explicit detail on these findings can be found in the Statistical Analysis Plan given to the client, however, the results will be briefly summarized here.

We found every variable in the client dataset to be a large departure from the general population through additional research and contrasting it with the sample collected through the firm using the methods of the design of experiments team. In light of these discoveries, we focused our analysis primarily on the data obtained by the consulting firm, which provides more insight into the USB Toaster product market. The data collected by the consulting firm is believed to be representative of the test market population that was proposed by the client for the USB Toaster. Given that this data sample was 100 times larger than that collected by the client and designed to be a random sample, we believe that the statistics gathered from this data set may be reliably used to infer conclusions about the population of these markets as a whole. We therefore used this dataset to achieve our specific aims enumerated previously.

Processing the Data

The original data obtained by the consulting firm included 45,211 observations with 11 potential variables of interest (Table 1). It was determined that a considerable amount of observations were missing at least some values in this dataset (Figure 1), with the variable *contact* having the most missing values at approximately 28% unknown or missing. For the purposes of data processing with categorical variables, unknown values and values with no entry were treated as one category. Across the categorical variables we tested proportionally against the population to see if there were any significant differences with respect to the outcome variable of product response. We found that there was no evidence that the missing values come from different distributions with the exception of *contact*; we did find that unknown values for *contact* were significantly different from the population and that this difference was somewhat large. The raw population mean for product response was 11.7%, however the product response for unknown *contact* was just 4.1%. Some further analysis of these missing values was conducted briefly using logistic regression and contingency tables, however no statistical or substantive reasoning for this difference was found by our team. Given that the number of unknowns for the *contact* category were so out of line with the rest of the data, and that the proportion of positive product response was so strongly affected by not knowing the *contact* type, we can see that this category is clearly meaningful for unknown reasons. It made sense to move forward with analysis in this case, but with caution in interpreting any results from the *contact* variable, however, recognizing that it remains an important part of the population as a whole and must not be ignored.

In every case of categorical variables the level unknown was kept as a factor level. We believe based on the proportionality test mentioned above that the missing values can be considered missing at random with respect to the product response in all variables except for *contact*.

Two continuous variables were also missing values: *age* (2586) and *balance* (795). In order to avoid listwise deletion of a significant proportion of the data, predictive mean matching was utilized to impute the missing values. Predictive Mean Matching (PMM) is a semi-parametric approach to imputation. This method is similar to the regression method except that for unknown values, a group of observed donor values whose regression-predicted values are closest to the regression-predicted value for the missing value are identified. Then using this group of observed donor values, the imputed value is selected randomly. This ensures that imputed values are plausible while also not relying on the assumption of normality for regression.

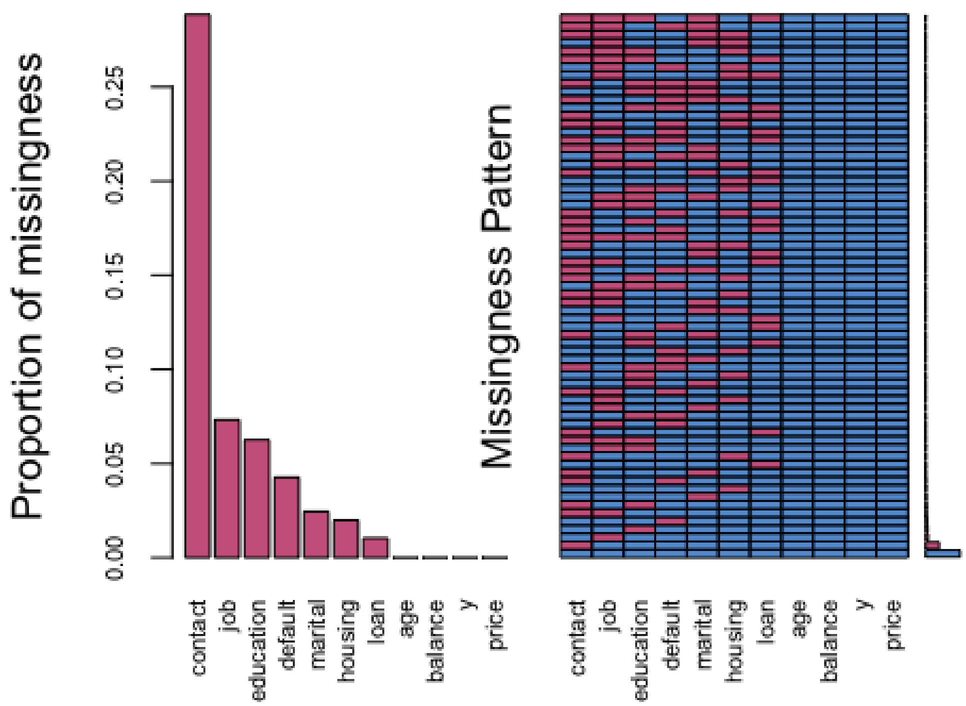
Table 1: Variable Table.

Variable	Type	Description
Age	Numeric	Age of the participant in the study.
Balance	Numeric	Non-mortgage loan balance.
Price	Numeric	The estimated price for the USB Toaster.
Product Response	Binary	Yes or no variable if the product response.
Contact	Categorical	Type of contact of the participant, cellphone, telephone, or unknown.
Loan	Binary	Yes or no variable if the participant has taken out a loan.
Housing	Binary	Yes or no variable if the participant has a mortgage.
Default	Binary	Yes or no variable if the participant's credit is in default.
Education	Categorical	Type of education of the participant, primary, secondary, or tertiary.
Marital	Categorical	Marital status of the participant, married, divorced, or single.
Job	Categorical	Type of job of the participant, blue collar, technician, management, administration, or services.

Five datasets of missing values were generated using PMM, each with 50 iterations. We chose to impute five datasets instead of one dataset in order to reduce bias. From the five datasets imputed, we selected the dataset with the distributions which best fit the previous distribution of the data.

Additionally, it was found that the variable *balance* had both negative and positive values, which was not reasonable given that this variable tracks the amount of non-mortgage loan balances. We believe this to be data entry errors since it would be reasonable that someone might construe debt as a negative value. Given this, balance was transformed by using the absolute value of the data found within the dataset.

Figure 1.: This shows the proportion of missing data by variable and the pattern of missing data respectively.



Note: This figure shows the data *after* age and balance have been imputed.

Data Analysis

SA1: How are the two datasets (from Table 1 and Table 2) different and why?

Although much of Specific Aim 1 (SA1) is addressed in the above data section, some more inspection of the client supplied data summary is warranted for informing further analysis done later. Although the differences in populations have been noted generally, there are specific features that are of interest. The client supplied data summary seems to be heavily skewed toward un-married, college males that may be considered less fiscally responsible, meaning that even if they wish to buy the USB Toaster they may not be able to.

It has been noted that surveys asking if people will hypothetically buy a product with money they may not have, is not necessarily a good measure for intent to buy. In researching the viability of this measure, our consulting team found a particular source of interest that specifically investigated durable homewares as part of its data. According to research published in the Journal of the American Statistical Association by F. Thomas Juster, "consumer intentions to buy are inefficient predictors of purchase rates because they do not provide accurate estimates of mean purchase probability." Thus, it may warrant caution when determining action based on this group.

The reasons for this uncertainty may be rooted in the fact that measuring an intent to purchase is a psychological construct and these measures have commonly known sources of unreliability, all of which apply in this context. The three common sources of performance variation are:

1) The person tested may change from one testing to the next.
2) The task measured and the behavior may be different.
3) The limited sample of the behavior results in an unstable and undependable score.

While it is not the intent to delve deeply into psychometrics, a brief aside exploring the potential reliability of the product response variable in the context of the research above might be fruitful for understanding the task at hand of examining the probability of product success. This will be explored further in the conclusion section.

As noted above, consumer intentions to buy have been found to not be the ideal measure for prediction of actual purchases. However, for our purposes we decided to take product responses at face value with the understanding that there is no way to untangle this potential misspecification within the data at present.

SA2: What is the likelihood, based on all available data, that the USB Toaster will be a success?

Using the data decided upon in SA1, we will focus on the average number of respondents that indicated they would purchase the USB Toaster. In the data set obtained by the firm, the percentage of respondents who indicated they would buy the USB Toaster via the product response variable was 13.25%, far short of the necessary 24.13% required to break even on the proposed product roll-out.

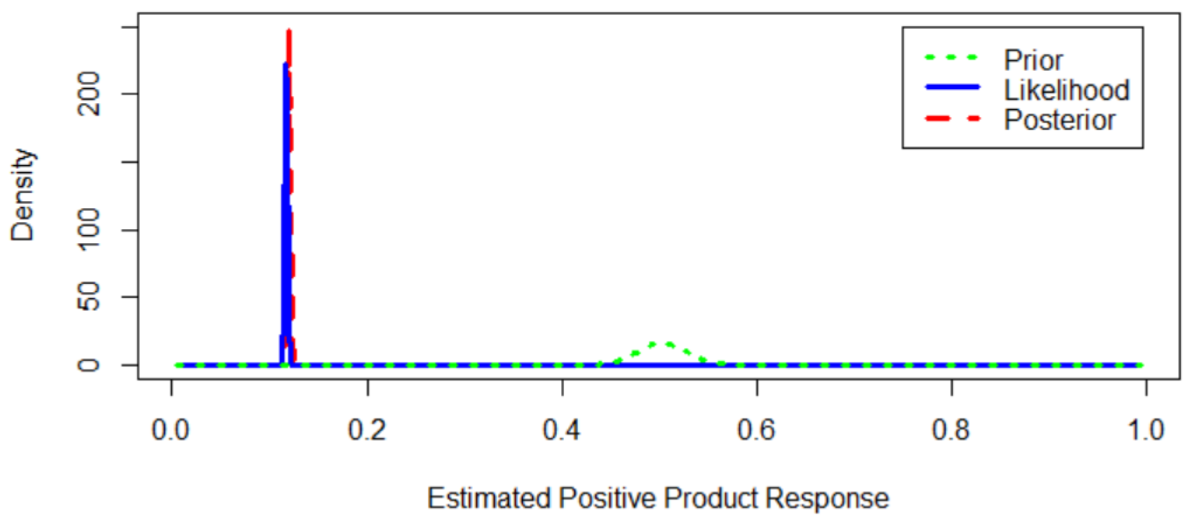
Following this, we conducted several potential intervals for the expected number of positive product responses, one each using frequentist methods for the data of the consulting firm and the combined data of the firm and the client, and others using Bayesian estimation that takes into account the previous information of the client dataset.

Table 2: Intervals of frequentist and bayesian approaches for estimating the mean.

Approach	Lower Bound	Upper Bound
Frequentist for data tables 1 and 2 (95%)	0.114	0.1199
Frequentist for data table 2 (95%)	0.1175	0.1235
Bayesian (95%)	0.1176	0.1236
Bayesian (99%)	0.1167	0.1245
Bayesian (99.9%)	0.1156	0.1256

Even allowing for the influence of the client dataset as a prior on the larger dataset gathered by our firm, we can see in Table 2 that the results are fairly robust. In Figure 2, we can see that the sheer size of the dataset gathered by the firm dwarfs the contribution of the mean supplied by the client. The end result being a strong confidence that at the given price and in the given market of the client, the USB Toaster will likely incur a substantial loss.

Figure 2.: The plot shows distributions of the expected value of probable positive products response based on prior belief of the client dataset (green), the likelihood given the consultant dataset (blue), and the posterior belief based on these two (red).



To explore this further, the probability of the product breaking even was calculated using the most favorable mean, the Bayesian estimated mean. Even using this mean, when calculating the probability of success using the beta posterior distribution, the probability was essentially zero that the USB Toaster would be successful in the target market.

SA3: What are some useful demographic data points that might help inform a potential marketing campaign?

To determine the answer to this question, we utilized two methods: logistic regression and contingency table analysis. We take each in part below and then interpret the results of both analyses afterwards.

Model Selection for Logistic Regression

In order to determine if there are potential demographics for which the USB Toaster might still be a profitable endeavor, we first built a logistic regression model using a training set of half of the data with product response as the outcome variable and all predictive variables mentioned earlier. By use of the Hosmer-Lemeshow statistic in the Pearson Chi-Squared distribution which uses (100) binned linear predictors as a measure, it was determined that the model was not of good fit and model selection began. For brevity, only the final model will be explored, though a description of the modeling process will be forthcoming.

The first step in model selection was using stepwise methods to determine which variables should potentially be included in the model, which did not result in a different model. When looking at the residuals by age, clear problems were detected and it was determined that typical demographic grouping by age might result in a better model. After grouping age by decade (i.e., 0 to 24, 25 to 34, 35 to 44, etc.) a new model (Table 3) was created which showed satisfactory fit according to the Hosmer-Lemeshow statistic for goodness-of-fit. While numerous additional models were considered using the Box-Cox transformations of *age* and/or *balance*, and log(*balance*), none were found to be of good fit. In Figure 3, we can see the fit of the 100 bins mentioned above in the Hosmer-Lemeshow statistic with bands of 95% confidence intervals to illustrate fit relative to the expected values for our selected model. Model interpretation will be conducted below.

The output of the model reveals that the variables of tertiary education, being employed as a maid, all age groups above 24, balance, yes for housing, that the participant has a loan, the contact methods of telephone and unknown, and marital status of married are the most significant to model the relationship.

Table 3: Coefficients and p-values for the logistic regression model.

Variable	Estimate	P(>|z|)
(Intercept)	-0.215	0.210
Education - Secondary	0.234	0.011
Education - Tertiary	0.486	**<0.001**
Education - Unknown	0.311	0.005
Job - Blue Collar	-0.251	0.005
Job - Entrepreneur	-0.479	0.003
Job - Maid	-0.646	**<0.001**
Job - Management	-0.175	0.049
Job - Retired	1.46e-03	0.991
Job - Self-employee	-0.176	0.189
Job - Services	-0.239	0.022
Job - Student	0.313	0.024
Job - Technician	-0.171	0.041
Job - Unemployed	0.065	0.628

Job - Unknown	-0.261	0.013
Age - 55-64	-1.117	**<2e-16**
Age - 45-54	-1.416	**<2e-16**
Age - 35-44	-1.306	**<2e-16**
Age - 25-34	-1.298	**<2e-16**
Age - <24	-0.542	0.006
Balance (in thousands)	2.13e-02	**<0.001**
Housing - Unknown	-0.337	0.035
Housing - Yes	-0.5621	**<2e-16**
Loan - Unknown	-2.56e-02	0.907
Loan - Yes	-0.477	**<0.001**
Default - Unknown	-2.92e-03	0.978
Default - Yes	-0.491	0.023
Contact - Telephone	-0.300	**<0.001**
Contact - Unknown	-1.550	**<2e-16**
Marital - Married	-0.254	**<0.001**
Marital - Single	9.32e-03	0.906
Marital - Unknown	6.34e-03	0.964

Note: P-values of significance 0.001 or less have been bolded.

Model Diagnostics

After determining the most likely model for the data, diagnostic analysis was performed. Residuals plots were constructed and examined. The residual plot for the binned linear predictors (η) reveals some curvature which was concerning (Figure 4). However, we were unable to resolve this with transformations and did not want to further hinder any interpretability, thus we examined the model further.

Figure 3.: This plot demonstrates the model's fit, where the values of the predicted probabilities align with the observed values. The farther from the line the points are, the worse the fit. The data points have 95% confidence intervals reflected in the lines, where most points at least fall in the 95% interval.

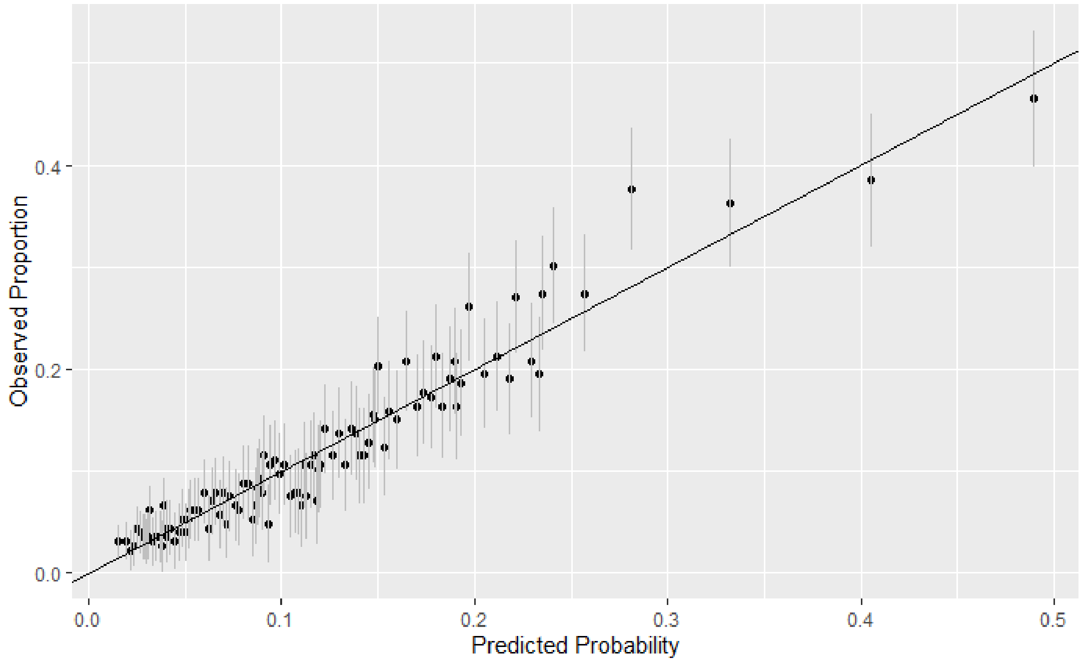
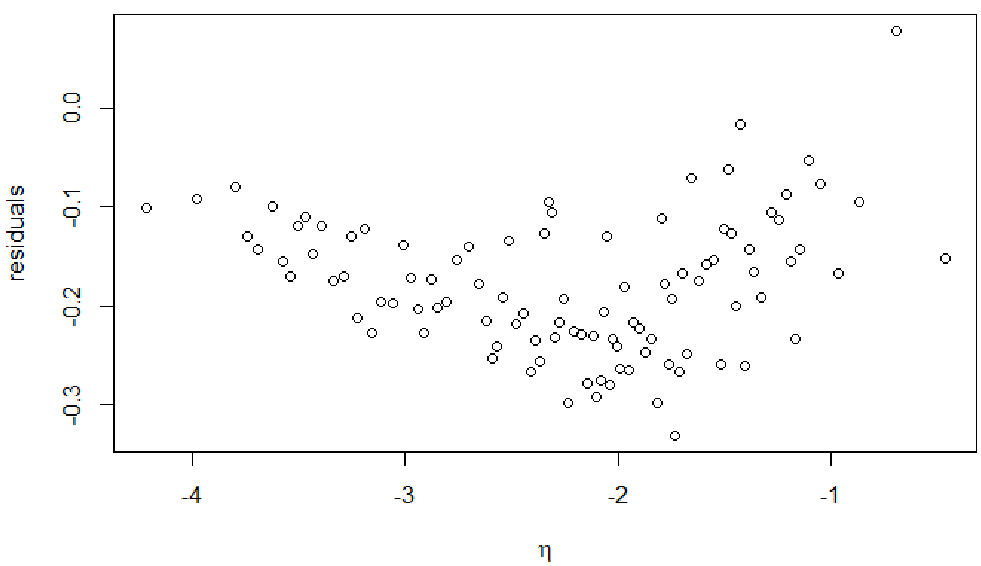


Figure 4.: The residual plot of binned linear predictors from the Hosmer-Lemeshow calculations for the final



model.

Next we considered residuals by individual variables. We found that certain values stuck out across the residual plots such as the age ranges of "younger than 24 years" and "65 years and up" and up in Figure 5 and "students" and "retired" categories in Figure 6. While many residuals were viewed, we have selected these two specifically because they fall into a larger pattern within our data and are consistent with the client summary data. Additional tests for leverage and various other concerns were performed with no results of note. These patterns will be examined in concert with additional analyses conducted via contingency tables below.

Figure 5.: The residual by age plot for the final model. Younger and older individuals behave differently.

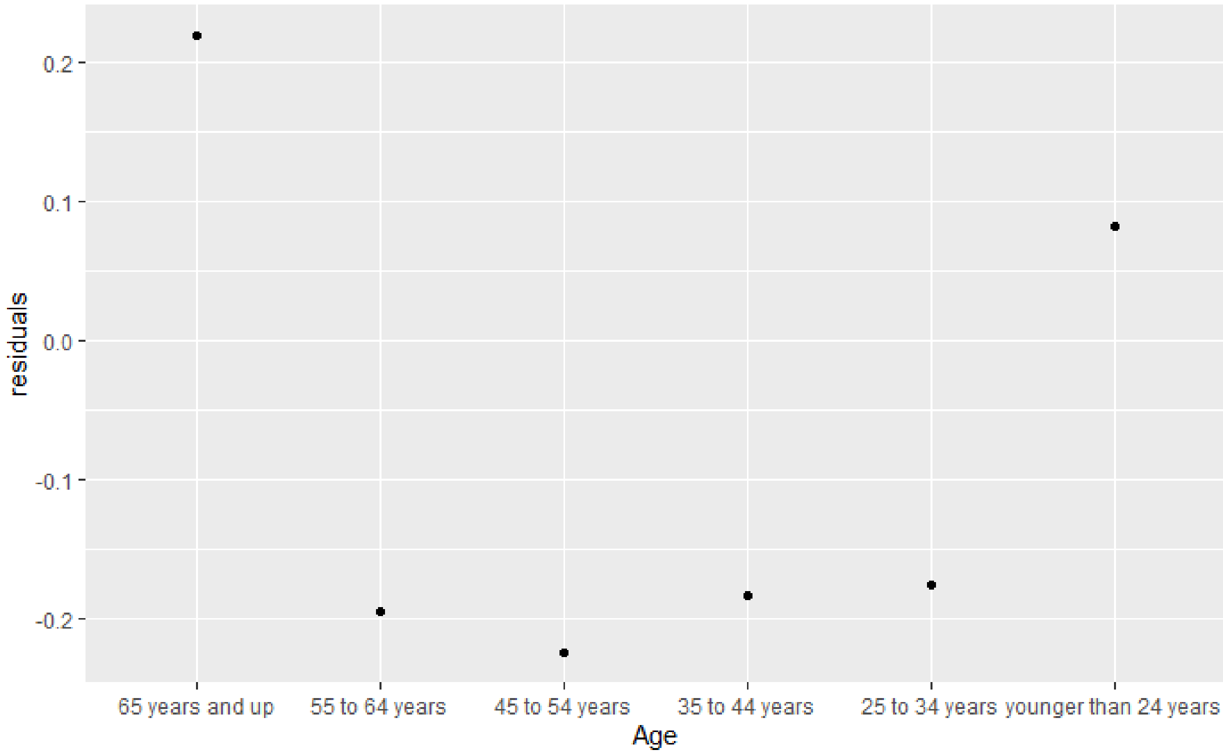
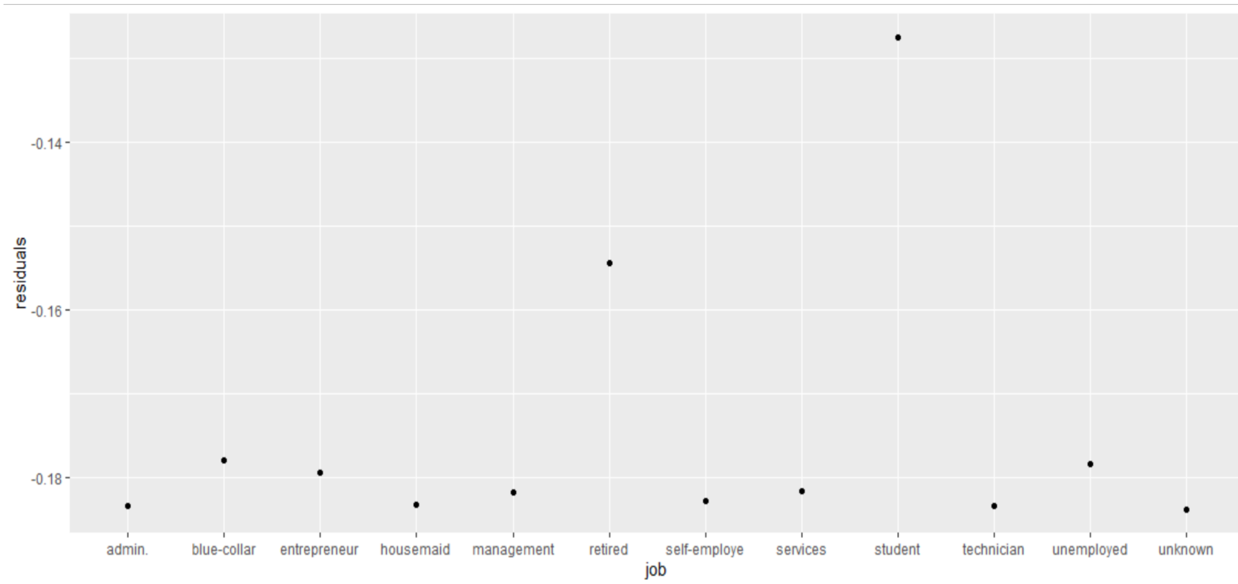
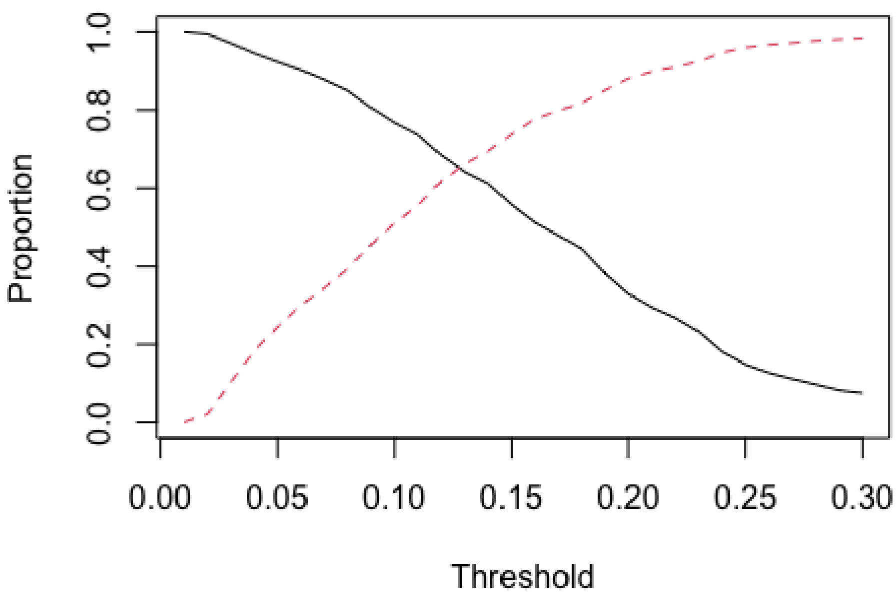


Figure 6.: The residual by job plot for the final model. Student and retired categories behave notably differently.



To further assess the model, the correct classification rate at a probability of 0.5 was calculated and determined to be 88.20%. Next the sensitivity (positive response correctly classified) was calculated and found to be less than 0.00%. We determined that the correct classification rate stemmed completely from the specificity (negative response correctly classified). Almost no cases are classified as "yes" because the probability of "no" always dominates the outcome, so the performance is not what we would like, but it is not surprising given the large proportion of "no" in each category. Thus, it is not coincidence that the correct classification rate is 88.2% when that is essentially equivalent to the negative product responses. In Figure 7, the sensitivity and specificity change can be seen for different thresholds of probability. While an optimal value exists, we do not believe it will be sufficient to consider the mold effectively predictive. This does not detract from the general directionality of correlations found in the model however. Given the lack of predictability, we did not check the test data set since the result was a forgone conclusion.

Figure 7.: A plot of the sensitivity (black) and specificity (red) of the final model for given probability thresholds.



Contingency Table Analysis

Contingency tables were constructed for all categorical variables both for those with positive product responses and the entire population. These tables of counts for positive responses were then divided by the table representing the count for total population. By doing this, we were able to see the percentage of each group of positive responses relative to the group at large. A few selected tables have been displayed in Tables 4 and 5, where we see distinct patterns of categories that exceed 20% positive product response. Following on the generated contingency tables, selective correspondence analysis for the frequency of positive product responses was conducted such as in Figure 8. In the correspondence analysis plot, central values are behaving according to the expectations of the model based on the information garnered from the first two diagonal values in singular value (SV) decomposition. The values of the diagonal contain the information from the matrix of residual values and appear in descending order, meaning that much of the information is carried in SV1 and SV2.

Table 4: Contingency table by percentage of positive product response for education and age.

Education/Age	<24	25-34	35-44	45-54	55-64	>65
Primary	**0.229**	0.069	0.064	0.057	0.088	**0.429**
Secondary	**0.235**	0.109	0.082	0.093	0.136	**0.430**
Tertiary	**0.230**	0.159	0.138	0.120	0.170	**0.393**
Unknown	**0.286**	0.131	0.115	0.099	0.121	**0.364**

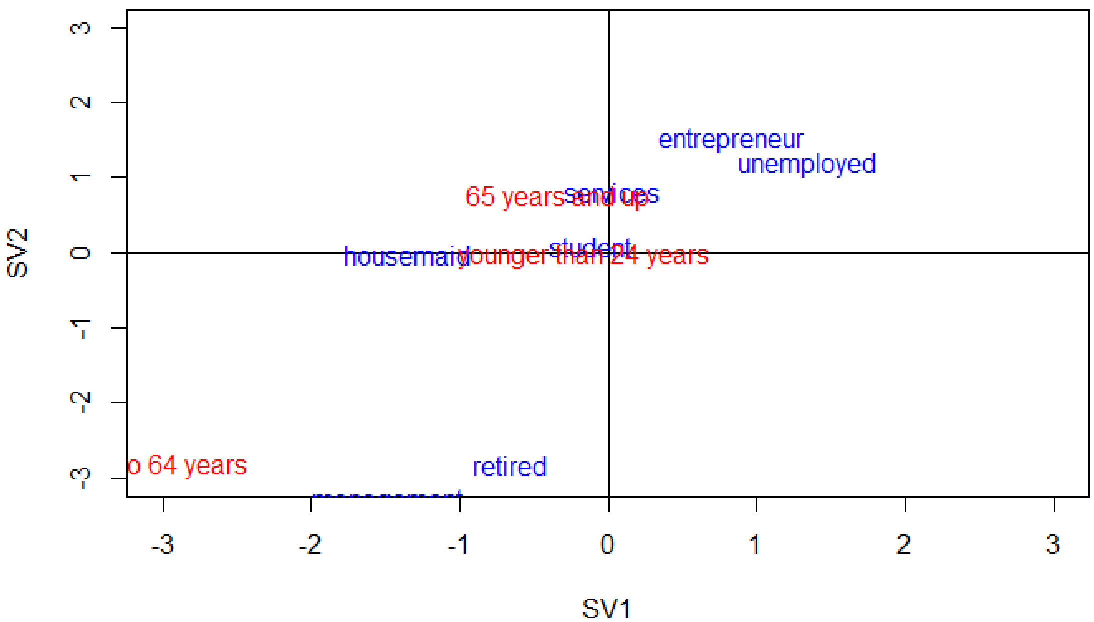
Note: Combinations with greater than 20% positive product response have been bolded.

Table 5: Contingency table by percentage of positive product response for job and age.

Job/Age	<24	25-34	35-44	45-54	55-64	>65
Admin	0.138	0.131	0.101	0.119	0.160	**0.500**
Blue-collar	0.152	0.085	0.063	0.067	0.690	0.000
Entrepreneur	**0.222**	0.075	0.088	0.081	0.040	**0.600**
Housemaid	**0.333**	0.076	0.069	0.080	0.098	**0.407**
Management	**0.368**	0.154	0.125	0.112	0.158	**0.341**
Retired	0.000	0.143	0.036	0.075	0.165	**0.415**
Self-employed	**0.400**	0.176	0.092	0.058	0.111	**0.667**
Services	0.165	0.096	0.085	0.076	0.079	NA
Student	**0.377**	**0.246**	**0.276**	**0.500**	NA	NA
Technician	0.159	0.119	0.107	0.098	0.135	**0.417**
Unemployed	**0.200**	0.144	0.154	0.145	**0.250**	**0.200**
Unknown	**0.220**	0.110	0.103	0.103	0.139	**0.457**

Note: Combinations with greater than 20% positive product response have been bolded. NA values are entered when no observations are present.

Figure 8.: A plot of the correspondence analysis of the singular value decomposition of age as a factor and job when considering positive product response. Central values tend to behave according to the model, showing that students and participants “younger than than 24” drive our model.



Combined Analysis

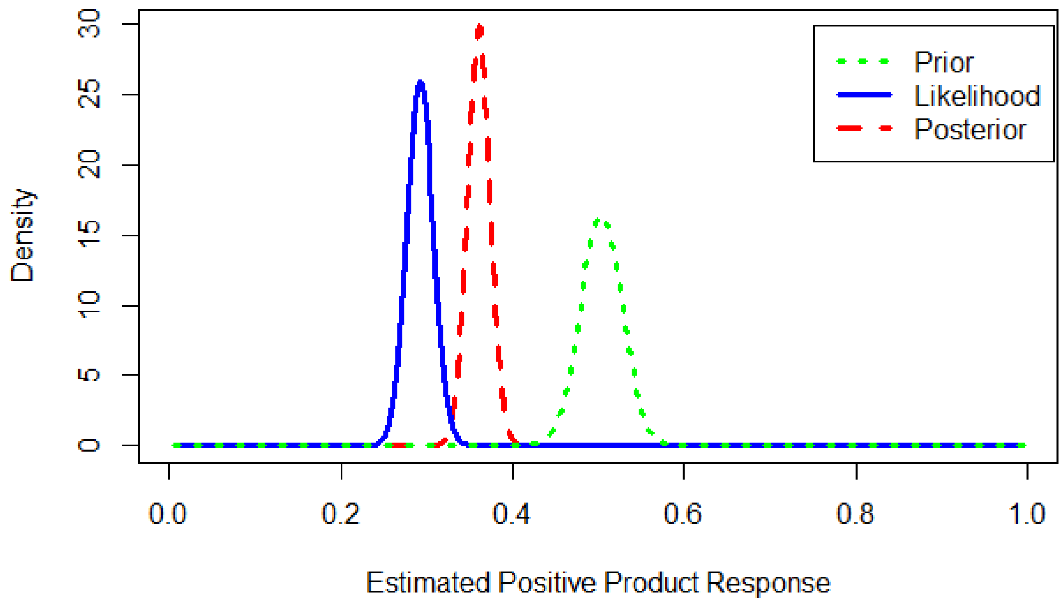
Across both the logistic regression model and analysis of contingency tables, we see a consistent story; individuals who are students or retired, with tertiary education, who are divorced or single, and do not have a house or debt are more likely to buy the USB Toaster. We see this both by virtue of positive coefficients in the logistic regression model and inflated count percentages in the contingency tables. Essentially, financially unencumbered persons who are well-educated and without families are the most likely to purchase the USB Toaster. Most striking was the student population, which was by far the most likely to indicate intent to purchase the product across the analyses. Following the results of both analyses new logistic models were briefly investigated (for the student, and student plus retired populations) but were determined not to be good fits.

SA4: What does the data tell us in terms of potential profit or loss based on the outcome of the analysis?

Based on the case of a probable loss in section elucidated in SA2, this analysis in SA3 may allow for stronger profit potential based on focusing on the demographics identified in the models that are positively associated with intent to purchase. Using the same Bayseian means estimation techniques as previously, with the client dataset as the prior, but now using only the student population, we have a much more encouraging outcome: 36.1% in the posterior within a credible interval of 33.5-38.7%. We also have the raw mean of the population of students at

29.2% with a confidence interval from 26.2-32.2%. We believe that the client sample may still be skewed positively because of the prevalence of males and other previously addressed abnormalities in the sample which might be more likely to buy. Although retirees and students were considered, after looking at the data more deeply, we felt only the student population size justified enough potential sales to be worthy of consideration as the target demographic.

Figure 9.: The plot shows distributions of the expected value of probable positive products response based on prior belief of the client dataset (green), the likelihood given the consultant dataset (blue), and the posterior belief based on these two (red).



It is worth considering whether the student population is large enough to achieve the same sales numbers as the five major markets originally targeted though. After some research, we determined that the population of the original market targeted was approximately 14.8 million, while the college student population of the United States is estimated at 18.9 million. Thus, the college market is sufficiently larger to actually have the potential to lower the necessary sales percentage for a breakeven point. Additionally, a significant portion of the five major markets are composed of children (28.5% of US households have children under 18) who would not be considered potential buyers, meaning that covering even 78% of the college market (equivalent to 14.8 million) would still exceed the expected market of the five major metropolitan areas. Since the original five metro areas also considered Toronto, Canada, this could expand the market even further than identified above. Understanding that these changes might influence the profit model, we address a few more potential sources of change below before a more detailed discussion in the conclusion.

The profit model given by the client is dependent on a 24.13% break even point and a fixed cost and price per item as well as fixed solicitation costs. We have proceeded with those as the anchoring points in our calculations from here forward. From this we can construct the profit for a population of size ‘n’ (in this case the 5 major metropolitan areas) to be

$$Profit_n = nz(\%purchase - 24.13)$$

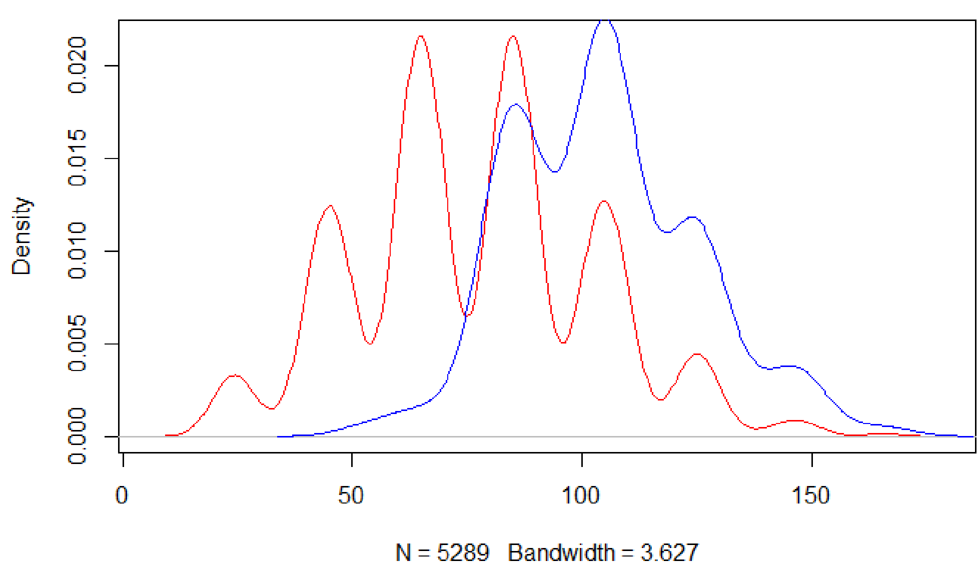
where z is the margin on one item (fixed price - fixed cost). This essentially gives us the profit per product sold times the difference of percentage of purchasers and the break-even point. Notably, this is for a particular fixed price for the USB toaster. While solicitation costs and cost per item might be truly considered fixed, it is possible that the client, based on estimated pricing data gained from the sample, might choose to raise the potential fixed price. In this case, raising the price of the product might lower the break-even sales point since

$$Breakeven * nz = solicitation\ costs$$

shows us that the break-even point and z are inversely related - meaning when margin per item goes up, the break-even point goes down. The same is true for n decreasing and the break-even point decreasing.

The pricing estimation information we've included may give additional insight into whether a change in price is justified based on the current proprietary price considered by the client. We will additionally assume the original population size, since this will allow the client to focus on reaching a percentage of colleges (78% or less) rather than all of them. We will be more than happy to run calculations and confidence intervals for both sales and profit should the client request this. For now, we will proceed with all fixed pricing and costs unchanged to see if the product is likely to be profitable as is.

Figure 10.: The plot shows distributions of the estimated price variable for the dataset population (red) and the student population (blue).



With regards to pricing, we can see that students who said they would buy the product expect on average to pay about $104.60, which is quite a bit more (roughly $28) than the population of yeses for the whole population ($76.79). Overall, more than 75% of the student population who would buy the USB Toaster expect to pay $87 or more with a minimum estimated price of $52. This pricing information might be crucial in estimating the profitability of the USB Toaster since any increase in price also alters the above z coefficient, increasing the per item margin. Thus, it might be worth considering an evaluation of the current proprietary price of the USB Toaster. In Figure 10, the distribution of positive responses for price is shown for the initial market (red) and the student market (blue).

Using the most conservative confidence interval of product sales from the student only population we will provide the calculation the 95% credible interval of profit or loss, keeping in mind this is set for a similar sized population and pricing as the original product roll-out, which might not reflect plans after changes based on this analysis.

Lower bound of profit:

$(391,230*z)

Mean Profit:

$(958,230*z)

Upper bound of profit:

$(1,525,230*z)

Conclusion

In the preceding sections, we set out to determine the feasibility of a profitable product launch for the USB Toaster. We determined that the roll-out plan as it is has an almost impossible uphill battle to profit. While this was an unfortunate outcome, we were able to identify demographics that responded more positively to the USB Toaster and find a potential market where it might succeed. Among the benefits of this new market are that it is highly focused and the client will know when and where students are likely to purchase. This is done by setting a launch date that might coincide with students moving to campus and focusing sales in locations such as college bookstores or college town department stores, that students are very likely to visit. We believe that the ability to target not only the potential customers, but places they are almost certain to go while already shopping is a clear benefit to the sales model. Additionally, by making small and low cost changes to the external appearance of the product, like putting a college logo on the exterior, we believe that the product will become even more enticing to this population and possibly improve sales to purchasers who might be on the fence.

Though the final analysis is indeed more rosy than the initial calculations on the outlook for the USB Toaster seemed, there is still considerable uncertainty involved. Based on the proprietary cost structure and unknown internal logistics for marketing and distribution, we were unable to determine whether this change in plans was feasible for the client. However, the significantly larger positive response on the part of this population will hopefully justify the changes and make the USB Toaster a success.

Still more uncertainty comes from the binary variable in product response. As mentioned above, there is the consideration of those potential buyers who might not be certain in either direction. In the article by Juster, this was the predominant source of error cited that made customer buying intentions problematic. As Juster puts it, “intentions surveys cannot detect movements in mean probability among non-intenders, who account for the bulk of actual purchases and for most of the time-series variance in purchase rates.” Meaning that though our data gives a “yes” or “no” response, the reality of the situation is different. This gets at the heart of each of the above sources of unreliability.

First, there is personal variation over time, whether through change in the person or the context of the person's life. Second, there is a difference in task - what we really want is the probability of a purchase, but the data forces a binary situation on the individuals. Finally, for each individual we have only one measure, which is not the most reliable measure. Juster’s article finds that having individuals rate the probability they will purchase a product is a better measure of predictability. While this does not completely fix any of the problems noted above, it allows for a better representation of the uncertainty in an individuals choice and thus makes for a better measure.

In future steps, it would be helpful to change the metric of the survey to be probabilistic to create a better understanding of an individual’s intent. Further steps include a different target demographic, specifically college students, as they appear to value the toaster at a higher price, as well as being more interested in purchasing it. Most students when moving into college are living alone for the first time, thus need to purchase necessary appliances. If the toaster was marketed in a way as a college dorm or apartment need, it could be more appealing to students.

Finally, as previously mentioned, the consulting team would be happy to do any additional work with the client with either more information about the proprietary price or manner in which the break-even point was calculated. Or alternatively, after the client adjusts their calculations of the break-even point based on the previous analysis to keep the information proprietary. We hope that the client is pleased with the analysis and the potential to turn a predicted loss into a likely success.