

# Technical Report

## **Investigation into the Potential Product: the USB Toaster**

Gokcen Buyukbas | gokbuyuk@indiana.edu

Minji Jeon | jeonmin@iu.edu

Luke Williams | willluke@iu.edu

### **Table of Contents**

1. Introduction
2. Analysis Population
3. Exploratory Data Analysis
3. Findings
4. Conclusion

## Abstract

The purpose of this exploratory study is to investigate the factors relevant to the purchase and the purchase price of the USB toaster. Additionally, this project was designed to compare the preliminary Sample 1 and Sample 2 collected from the Design of Experiments team (DOE), and to justify the collection of Sample 2. Sample 2 was a well-designed sample: the DOE sampled randomly from the five major metropolitan areas targeted for the initial product roll-out. We explored the factors (explained in Table 2) relevant to purchase and purchase price by applying a multiple regression analysis on Sample 2 data. To analyze the data properly, we first identified and agreed upon a process to address missing and incomplete observations. From the logistic regression model on *purchase*, we found that the relevant factors were *age*, *profession*, *marital status*, *education*, whether they have a *mortgage or personal loans*, *non-mortgage loan balance*, and types of *contact device*. From the regression model on *price*, we found similar results of significant factors: *profession*, *education*, *marital status*, whether they have a *mortgage or personal loan*, and type of *contact device*, but we also found *credit default* and the *purchase* of the product had a significant impact as well. From this information, there is a better understanding of potential target audience for the USB toaster to maximize sales and profit.

## 1. Introduction

With the increase in access to computers, many different occupations have changed how they conduct business. Often, finding a warm quality breakfast or lunch is difficult in the office. People often miss having a warm piece of toast with their morning coffee because of the inconveniences of getting it. That is where the USB Toaster comes into play. The USB Toaster is a new product in development that attaches to your computer and provides you with a quality meal wherever you may be. Because of the small size of the USB toaster, it is easy to transport, and it does not require an outlet.

The USB toaster is a new product, therefore there is not much research into its target audience or the optimal price. The USB toaster will be ready for the stores following product and marketing testing. The purpose of this document is to report the results of investigating the demographics of the target audience in order to maximize sales and profit. Additionally, this project was designed to compare the preliminary Sample 1, and Sample 2 collected from the Design of Experiments team (DOE), and to justify the collection of Sample 2. Sample 1 had a very small sample size and was not representative of the population, while Sample 2 is much larger, and it was a random sample taken from the population. Further analysis will be discussed in the analysis population.

In the product analysis we looked at two different regression models based on the binomial response of whether the customer would purchase the product, as well as the response of suggested retail price. From the logistic regression model on *purchase* (hereafter, Purchase Model), we found

that the relevant factors were *age*, *profession*, *marital status*, *education*, whether they have *mortgage or personal loans*, and types of *contact device*. From the regression model on *price* (hereafter, Price Model), we found important factors of *profession*, *education*, *marital status*, whether they have a *mortgage or personal loan*, type of *contact device*, *credit default*, and the *purchase* of the product in the prediction of price.

In this paper, we will start with a preliminary analysis by comparing the quantitative and qualitative differences of two samples and justify the expense of the large sampling conducted by the DOE. Following this we will investigate two different regression models, logistic regression for *purchase* (Purchase Model), and multilinear regression for *price* (Price Model) in the modeling process. In the last section we will discuss the study and conclude the paper by providing the most important results from this analysis and propose the optimal marketing strategy and price as well as an estimated profit of the USB toaster considering our analysis.

## **2. Analysis Population**

The USB toaster is designed for the general population of adults. Two surveys were conducted and the demographic profile of the two samples are summarized in Table 1. The initial data was collected from an online direct marketing campaign conducted by the firm. Preliminary analysis of this campaign indicated that the product would be lucrative based on initial results showing that 50.36% of the sample responded favorably to the product. Internal accounting records estimate that the break-even response rate is 24.13% based on sales revenue, production costs, and solicitation costs. However, this finding should be taken with a caveat considering whether the sample in the initial survey represents the potential customer population for the product. The initial dataset appears to have sampling problems in several demographic profiles. For example, the average age of Sample 1 is 31.5 years old, which shows that the initial data was sampled from a younger population (see Table 1). Furthermore, 76% of them were male, and 86% of them were white. One glaring example of this biased sample can be seen in the professions of the respondents, 65% of whom are students. Therefore, it calls for another survey that more accurately displays the possible consumer population. The second study covers a larger subset and a better representation of the population eliminating the biases described above.

**Table 1.***Demographic Summary for Sample 1 and Sample 2 Grouped by Purchase*

	<b>Sample 1</b>			<b>Sample 2</b>		
<b>Variable</b>	<b>Yes</b>	<b>No</b>	<b>Total</b>	<b>Yes</b>	<b>No</b>	<b>Total</b>
Number	212	209	421	5,289	39,922	45,211
Gender (Male)	82.0%	71.0%	<b>76.0%</b>	NA	NA	NA
Race (White)	89.0%	82.0%	<b>86.0%</b>	NA	NA	NA
Age (Years)	25.5	37.5	<b>31.5</b>	41.6	40.8	40.9
Marital status (Married)	<b>1.0%</b>	27.0%	14.0%	50.4%	59.8%	58.7%
Education (College or more)	93.0%	73.0%	<b>83.0%</b>	36.9%	27.7%	28.8%
Education (Secondary)	44.0%	51.9%	50.1%	45.3%	50.7%	50.1%
Mortgage (Yes)	<b>1.0%</b>	10.0%	5.6%	35.8%	57.0%	54.5%
Personal loan (Yes)	6.3%	15.6%	13.5%	9.1%	16.8%	15.9%
60+ day delinquency	69%	9%	44%	NA	NA	NA
FICO	NA*	620	620	NA	NA	NA
Primary phone (Cell)	91.9%	91.2%	91.4%	82.5%	62.4%	64.7%
Profession						
Administrator	NA	NA	NA	11.0%	10.6%	10.6%
Management	NA	NA	NA	19.0%	22.8%	19.5%
White Collar	20.0%	5.9%	13.0%	NA	NA	NA
Blue collar	3.0%	1.0%	2.0%	12.5%	21.1%	20.0%
Entrepreneur	NA	NA	NA	3.2%	2.1%	3.1%
Retired	NA	NA	NA	9.0%	4.1%	4.7%
Self-employed	NA	NA	NA	3.3%	3.3%	3.3%
Services	NA	NA	NA	6.5%	8.9%	8.6%
Student	67.0%	64.0%	<b>65.0%</b>	4.8%	1.6%	<b>1.9%</b>
Housemaid	1.0%	29.0%	15%	1.9%	2.6%	2.6%
Technician	NA	NA	NA	15.1%	15.8%	15.7%
Unemployed/unknown	9.0%	0.3%	4.7%	4.1%	3.2%	3.3%
Loan balance (\$)	23,879	1,250	12,645	1,807	1,307	1,365
Contact duration (in seconds)	404	215	259	533	221	258
Number of contacts	1.8	2.1	2.1	2.1	2.9	2.8
Days after contact	193	234	225	68.8	36.4	40.2
Previous contacts	3.2	3.2	3.2	1.2	0.5	0.6
Price (\$)	82.2	42.6	51.8	76.8	48.8	52.0

*Note:* Mean is displayed for the continuous variables, percentages are displayed for categorical variables. Most respondents did not have a FICO due to “thin” credit.

In Sample 2, the extensive data collected from the DOE was taken from a well-designed random sample from five major metropolitan areas (Toronto, ON; New York, NY; Philadelphia, PA;

Dallas, TX; and San Francisco, CA) for the initial product rollout. The sample size of Sample 2 is large at 45,211 participants, compared with the sample size of 421 in Sample 1. In addition, Sample 2 presents a more representative sample, as it reflects a more comprehensive age distribution and reasonable proportions of various professions, and reduced differences of loan balances between the positive and negative answers (see Table 1). Accordingly, we conducted analysis on the current dataset investigating how the willingness to purchase the product varies by a consumer's demographic profile based on Sample 2. The potential variables of interest include age, profession, marital status, education level, credit, and loans, which are described in Table 2.

**Table 2.**  
*Descriptions of Variables*

Variable Name	Description
Age	Integer; In Years
Profession	Categorical: administrator., unknown, unemployed, management, housemaid, entrepreneur, student, blue-collar, self-employed, retired, technician, services
Marital status	Categorical: married, divorced (or widowed), single
Education	Categorical: unknown, secondary, primary, tertiary
Credit default	Binary(Y/N); has credit in default?
Non-mortgage loan balance	Non-mortgage loan balance
Mortgage	Binary(Y/N); has mortgage?
Personal Loan	Binary(Y/N); has personal loan?
Primary phone	Categorical: unknown, telephone, cellular
Contact duration	Last contact duration; In seconds
Number of contacts	Number of contacts performed during the campaign and for this contact
Days after contact	Number of days that passed by after the contact was last contacted from a previous campaign (-1 means they weren't previously contacted)
Previous contacts	Number of contacts performed before this campaign and for this contact
Previous outcome	Outcome of the precious marketing campaign. Categorical: unknown, other, failure, success
Purchase	Binary(Y/N); has the contact indicated they will purchase the product?
Price	Suggested/being willing to pay retail price, In USD

### 3. Exploratory Data Analysis

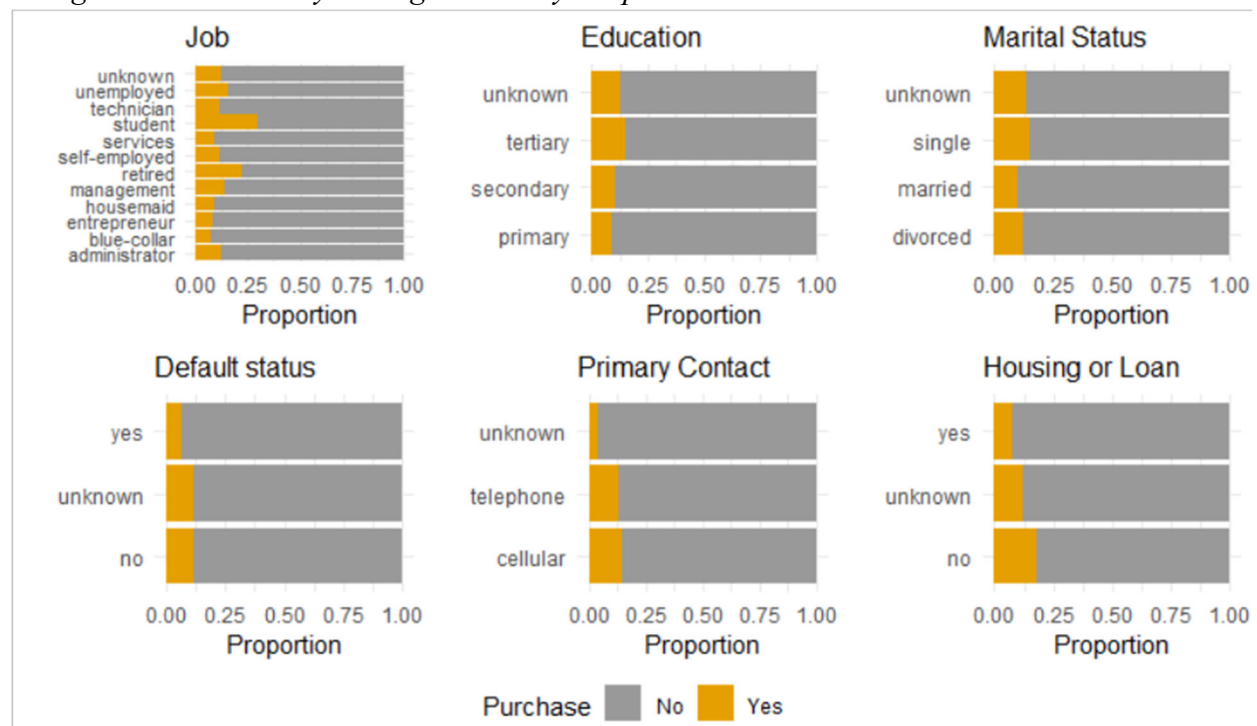
We started our analysis with an exploration of Sample 2. We excluded the day and month as we thought they were irrelevant to our market research. There was a substantial proportion of missing values, which is present with any survey, with 46% of the rows having some missing values additional to those that were unknown. We converted missing values to unknown for the categorical variables and treated them as a different level. This reduced the rows with missing

values to 21%. These were the missing values in the numerical variables; we removed them from and that left us 36134 observations to work with.

The proportions of those who are interested in buying the USB Toaster can be seen in Figure 1 for each categorical variable. Students and retired people seem to be the most interested in this product. As the education level increases, the interest in the product also increases based on our initial exploration of the proportions in Sample 2. Single people seem to be more likely to buy the USB Toaster as one could expect. Those who have credit in default or have a housing or personal loan are less likely to buy the product. We also observe that cellphone users as their primary contact are more likely to buy a USB Toaster.

**Figure 1.**

*Categorical variables by willingness to buy the product*



The distribution of the numerical variables is similar for those who are interested in the product and those who are not interested. We see some difference in the duration of the last contact, the call was longer for those who are interested in the product. This is not surprising, since if they are interested in buying the product, they would want to hear more about it. Also, as the number of days since last contact increases, the number of customers who are interested in the product decreases.

**Figure 2.**

*Numerical variables by willingness to buy the product*



*Note: All variables except Age are in logarithmic scale*

In Figure 3, we see a right skewed distribution for the suggested price by those who are not interested in buying the product with a mode at \$45. However, the distribution of the price suggested by those who are interested in buying the toaster seems to be multimodal. This may be due to the income levels of different professions of those who want the product.

**Figure 3.**

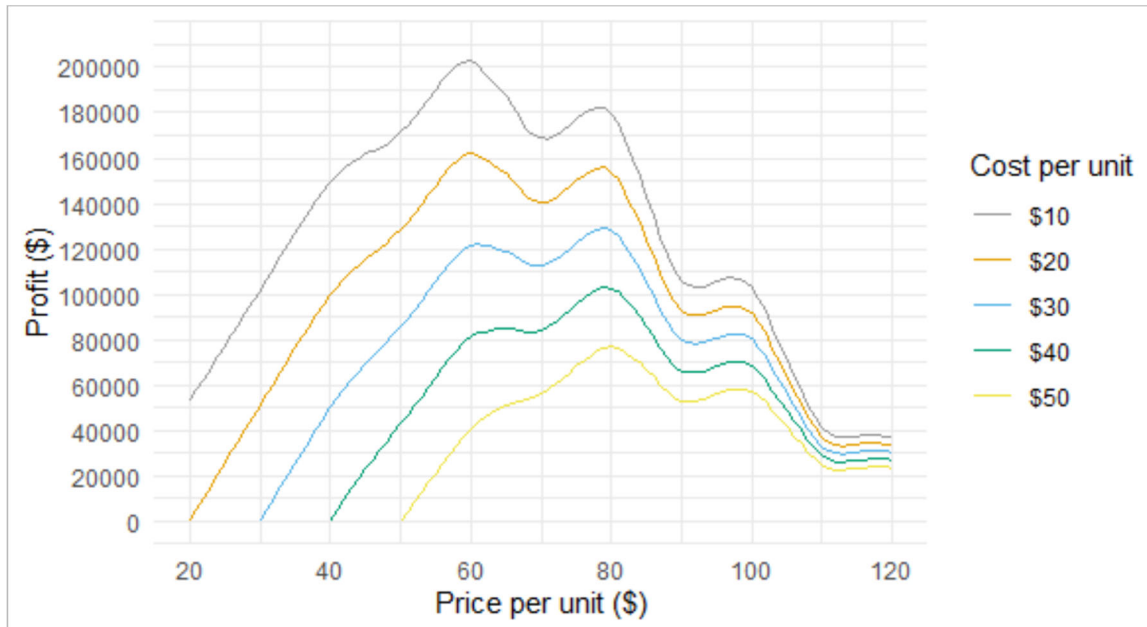
*The distribution of the suggested price by the willingness to purchase.*



For pricing, we calculated the profit based on different costs per unit assuming that our customers are only those in Sample 2. As seen in Figure 3, the best price for a broader target audience is \$60 per unit when the cost is below \$30. As the cost per unit increases, a price of \$79 per unit becomes more profitable for the broad audience. We see another potential price of \$97 per unit that might be profitable if we target a specific audience.

**Figure 4.**

*Estimated profit based on several levels of marginal cost per unit.*



#### 4. Findings

From our two models, Purchase Model and Price Model, we intended to reveal the target customer profile and predict the price that maximizes the profit.

##### 4-1. What would be the demographic profile of the target consumers?

From the logistic regression analysis for Purchase Model, we found the relevant factors are age, profession, marital status, education, non-mortgage loan balance, whether they have a mortgage or personal loans, and types of contact device (see Table 3). The retired and students are more likely to buy the USB toaster than those with other professions. Particularly, the probability these groups of people would buy the product increases by 172% (i.e., 2.72 times the odds) compared with entrepreneurs and housemaids. Another relevant factor, age was shown that one more year leads to an increase in a purchase by 0.5%. Unmarried singles are willing to buy the USB toaster than the married or divorced (and widowed). The more highly educated, the more responded to make a purchase. Those who had mortgages or personal loans were less inclined to buy the product by 50%. The respondents who were using a cell phone responded more favorably than those with a telephone.



**Table 3.***Demographic Profiles Influencing the Purchases*

<b>Variables</b>	<b>Coefficient Estimate</b>	<b>Probability Increase in Purchase (%)</b>	<b>P-value</b>
<b>Age</b>	0.01	1.01	<.01
<b>Jobs</b>			
Student	0.56	75.07	<.001
Retired	0.47	60.00	<.001
Entrepreneur	-0.49	-38.74	<.001
Housemaid	-0.42	-34.30	<.001
<b>Marital</b>			
Married	-0.18	-16.47	<.001
Single	0.22	24.61	<.001
<b>Education</b>			
Secondary	0.18	19.72	<.01
Tertiary	0.38	46.23	<.001
<b>Mortgage or Personal Loan</b>	-0.70	-50.34	<.001
<b>Contact: Telephone</b>	-0.26	-22.89	<.001

**4-2. What is the right price and how can we maximize the profit?**

Significant variables of interest in the regression model on Price can be found in Table 4. The estimate is obtained from transformations from the log Price model (to linear) so that the estimates directly relate to *price*. The log-odds estimate is obtained from the model. (Coefficient Estimate is in US Dollars)

**Table 4.***Price Model Results*

<b>Coefficients</b>	<b>Coefficient Estimate (\$)</b>	<b>Probability Increase in Price (%)</b>	<b>P-value</b>
Profession - Student	20.4	56.99	<.001
Education -Tertiary	1.9	4.39	<.001
Marital Status - Single	18.9	54.96	<.001
Default - Yes	18.1	52.95	<.001
Contact - Telephone	2.1	4.50	<.005
Purchase - No	-23.8	-37.99	<.001

Initially, the response variable of price was looked at, and based on the exploratory data analysis, it was found that a log transformation would best fit the shape of the Price of the USB toaster. The linear price model did a fairly good job with the data, but a log-transformation fits better with the distribution of the variable *price*. After selection of the log-transformed multiple regression model, the variables of interest were found to be the *profession*, *education*, *marital status*, whether they have a *mortgage or personal loan*, type of *contact* device, *credit default*, and the *purchase* of the product in the prediction of price. Each variable of interest had a strong association with the price. The category of *profession* that had the greatest difference in log-odds was students, who had a \$20.4 increase in the average retail price of the product. In the model, it was found that a higher level of education increased the retail price that would be paid. Those who were listed as single and those who defaulted on their loan both had close to an 18 increase in price. The people that responded to the contact as unknown had a log-odds of .409 (50.53% increase) which is a remarkably high increase in odds for the price, but since it provides no information relevant to the demographic target audience, it is better to look at the next best contact which was the telephone contact. Lastly, those who did not want to purchase the product valued the product \$23.8 less than those who wanted to buy the product. Looking at the target demographic audience (excluding credit default and telephone response- since those aren't demographics), the median purchase retail price is \$97, with a 95% CI (Confidence Interval) from \$91 to \$103 for the true median. Those who responded that they wanted to buy the product, and had these demographics assigned the retail at a much greater value of \$121. Both price estimates are much greater than the median value for the entire population of \$50.

## 5. Conclusion

The purpose of this study is to investigate the demographics of the target audience to maximize profit and support the additional sample conducted by the DOE. Overall, Sample 1 had a small sample size and was not representative of the population, while Sample 2 is much larger, and it was a random sample taken of the population so it will better help determine target audience and demographics as well as price.

Based on the Purchase Model looking at the demographics of who should the USB toaster be marketed to, we found the following factors to be significant: the retired or students, unmarried singles, having undergone tertiary or secondary educations, without mortgage or personal loans, and responding with a cell phone. Looking at the Price Model, it was found that the desired audience for the product is students in advanced schooling, that are single. Those who are students with advanced degrees and are single are the group of people that have access to their computers 24 hours a day, so the convenience of the USB toaster would be a big advantage for them. Additionally, those who are students in higher education and single are most likely to have a busy schedule and no time to cook for themselves and don't have a family member cook breakfast for them. That creates an added need for a good breakfast on the go without using money to go to a

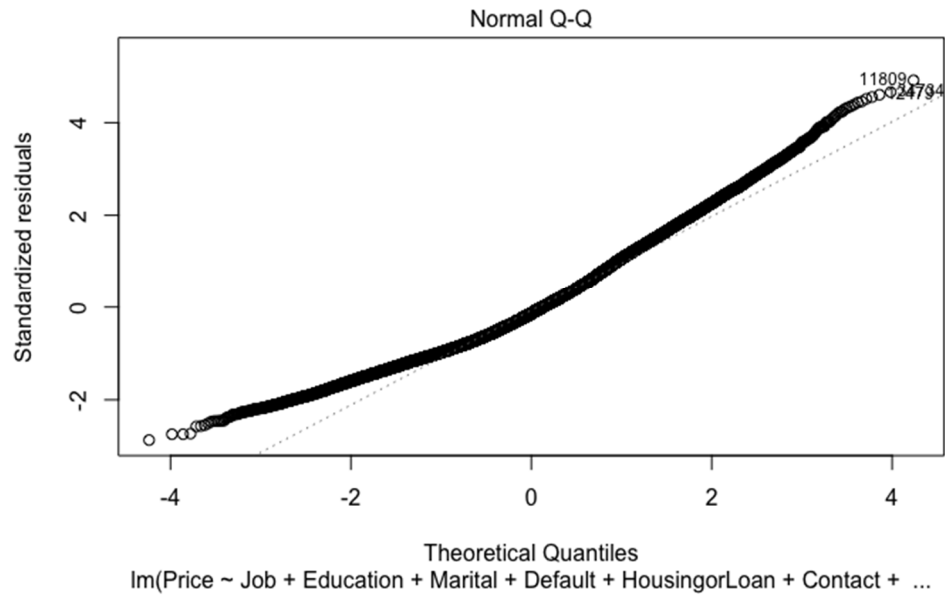
quick breakfast place, which is a rationale for this audience valuing the USB toaster much higher. The respondents were also more likely to have a higher price if they had defaulted on their loan and they contacted through telephone, although those that contacted with a telephone were less likely to buy the product. Both variables comment on the socio-economic status of those who want to buy the product, as they are both correlated with those who are in lower economic classes. Although it may seem like they would value the product less than richer people, the lower classes are more likely to need the product because the upper classes can get access to a normal toast at home or at work. Upper classes might have less use for the product, which could be why they value it less.

Without any specific target audience, we recommend a price of \$60 if the cost per unit is below \$30. If the cost per unit is between \$30-\$50, then \$79 is a more profitable price for the broad audience. Based on the demographics of the target customer that maximizes the profit, the median price is around \$97, which is a high estimate, since this is the most profitable target audience. The price could be estimated with a better knowledge of other factors such as product cost and supply of USB toasters. To better understand the optimal price to maximize the profit, it may be useful to run optimization once the USB Toaster is tested for cost, supply, and demand. To determine the price and target audience better, it may be beneficial to design an experiment so that we could determine the causal relationships with the price of the USB Toaster and profit.

## Appendices

### Appendix A

#### Linear Price Model Quantile-Quantile Plot



### Appendix B

#### Log Price Model Residual v Leverage Plot

