ANLY 533 – PCA and Logistic Regression

Improving mailing performance by predicting consumer likelihood to purchase a service contract.

Analyst: Adam Sampson

Contents

[Introduction and Business Problem 2](#_Toc498290128)

[Assumptions and Methodology 2](#_Toc498290129)

[Results 2](#_Toc498290130)

[Brand Effects: 2](#_Toc498290131)

[Product Group Effects: 2](#_Toc498290132)

[MSRP Effects: 3](#_Toc498290133)

[Offer Price Effects: 3](#_Toc498290134)

[Census Data Effects: 3](#_Toc498290135)

[Performance Curves 3](#_Toc498290136)

[Recommendations: 3](#_Toc498290137)

[Appendix A: Logistic Regression Model 4](#_Toc498290138)

[Appendix B: Odds Ratios 5](#_Toc498290139)

[Appendix C: PCA Top 10 Contributors by Component 5](#_Toc498290140)

# Introduction and Business Problem

Mailers for service contracts on appliances are currently resulting in 3.04% service contract purchase rate. Service contracts are worth either $60 or $70 revenue, less the expected payout on a service contract. Low conversion rates are detrimental to the marketing campaign. At this time, not enough information has been provided to determine the profitability of the current or future campaigns.

The objective of this analysis is to explain a consumer’s likelihood to purchase a new service contract. This data will need to be combined with campaign cost and profit expectation numbers to determine path forward.

# Assumptions and Methodology

The method used to perform the analysis is included in the SampsonRegressionAnalysis.R script. For further information look there.

In brief, a Principal Component Analysis was applied to provided census data for the block group of each customer. This was combined with the sales, brand, model, and pricing data to perform a logistic regression analysis.

# Results

The results of the analysis were a model which predicts the purchase of a service contract using Brand, Product Group, MSRP of the appliance, Offer Price of the service contract, and 5 principal components of the census data.

It is important to note that all values in the regression are from the base variable. Brand results are compared to Average Brand A. Product Group variables are compared to Washers. Offer Price is compared to a service contract of $60. To see how much more likely someone is to buy a contract use the odds ratio in Appendix B.

### Brand Effects:

Likelihood of buying a service contract relative to reference Average Brand A:

* + 16.9% for Premium Brand A
* + 5.4% for Average Brand B
* + 2.2% for Premium Brand B

### Product Group Effects:

Likelihood of buying a service contract relative to reference product group Washing Machines:

* +15% for Ice Makers
* -28% for Stacked Laundry
* -30% for Freezer
* … etc, see Appendix B
* -65% for Cooktop
* -67% for Range Hood
* -74% for Air Conditioners

### MSRP Effects:

MSRP was not a major effect in determining the odds of a sale.

### Offer Price Effects:

Likelihood of buying a $70 contract was 11% higher than that of buying a $60 contract.

### Census Data Effects:

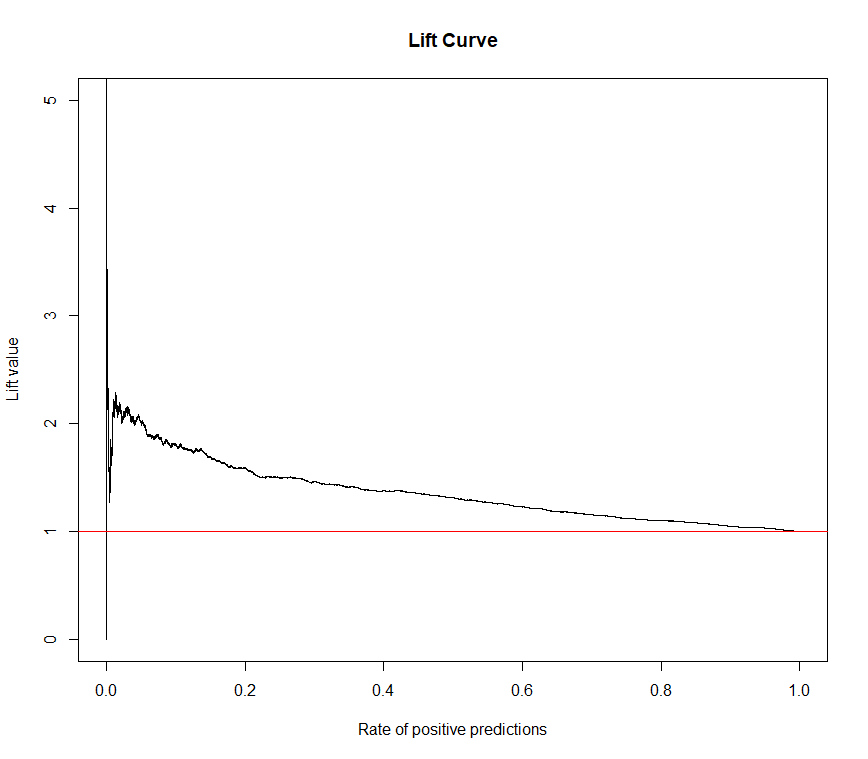
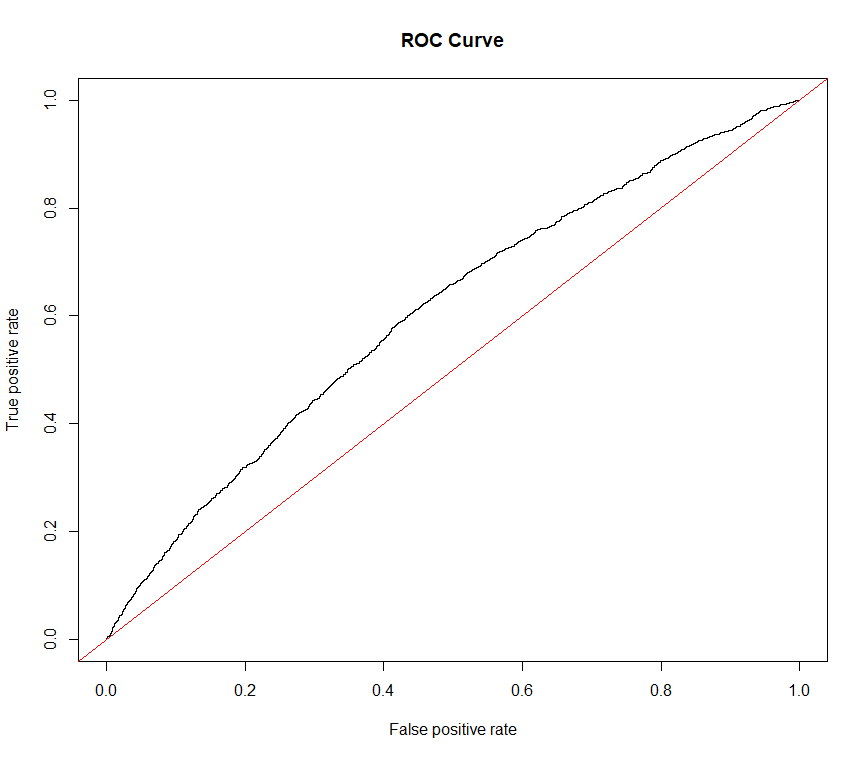
A unit change in principal component 2 (PC2) had a +22% effect on likelihood of buying a contract. The other principal components had a much lower impact per unit change. PC5 had the second most with 5% effect per unit change. PC1,3,4 had less than 2.7% effect on the odds per unit change. The top 10 features in each component are located in Appendix C.

PC2 is categorized by a customer being from a census block with:

* Less households (Family households and Single Households were both negative)
* Less white people
* Less units in detached buildings
* Less expensive rent prices (over $1000)
* Less cheap rent prices (less than $500 or less than $200)
* Less married people
* Lower percentage of men

In other words PC2 has a lot of census block areas that are more rural, have higher than average percentages of minorities, or have more single women.

### Performance Curves



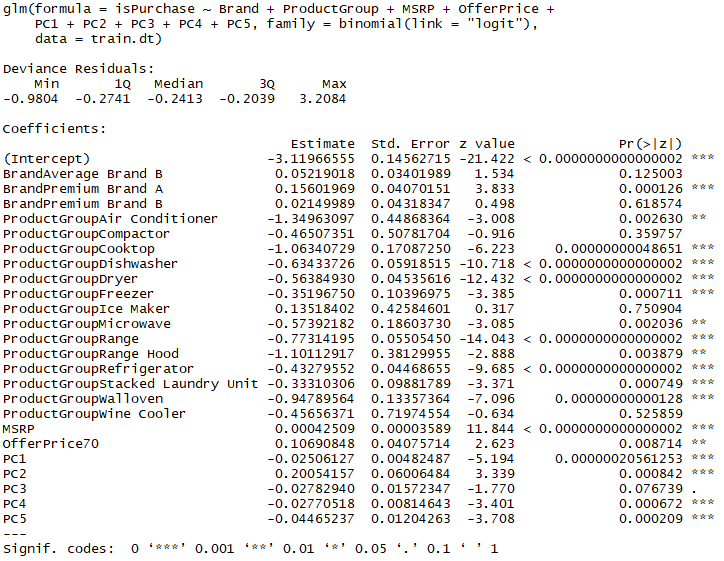
Area Under the Curve is 60.7% which indicates the probability that a positive sale value will be ranked before a negative/failed sale value. This value is poor, but higher than fail.

# Recommendations:

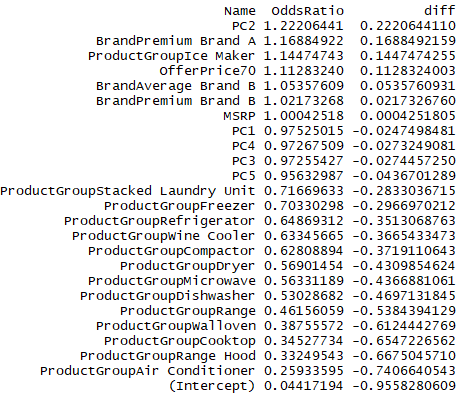
This model is useful for generating predictions of the probability that a mailing will result in a sale. Depending on the number of mailers that are going to be sent out, it generates lift of between 1 and 2. Of the data included, the gains begin to fall off at slightly less than 50% of the mailers that were sent out.

Further analysis could be performed to break the model groups into more specific product details. However, this would require time investment to clean up those data fields.

# Appendix A: Logistic Regression Model



# Appendix B: Odds Ratios



# Appendix C: PCA Top 10 Contributors by Component

