Prediction of Customer Purchase Behavior

PROJECT REPORT FOR MSIA 401: PREDICTIVE ANALYTICS (I)

MENGSHAN JIN
LINGXIAO OU
APURVAA SUBRAMANIAM
LUYAO YANG

NORTHWESTERN UNIVERSITY

DECEMBER 4, 2015

Executive Summary

This report provides an an analysis of customers' purchase behaviors during our Fall promotions. Specifically, we used statistical methods to predict whether customers will make purchases, and if so, their purchase amount.

Our models are able to correctly identify the purchasing decisions of 93.5% of the customers, and allow us to estimate their spending amounts with an average error of $7.30. We estimated a revenue of $174,238 among the best 1,000 customers, which is close to the actual total of $165,965. Among the top 1,000 customers chosen by our model, 443 actually made a purchase. This resulted in a revenue of $57,432 for the top customers. Comparison of different techniques, as well as graphics analyzing general trends within our data, can be found in the appendices.

Analysis shows that the most valuable customers are those who:

- are frequent buyers of the company's products,
- have made a purchase during the past promotions, and
- tend to spend more per order.

The prediction models can be further improved by collecting data on the following items:

- Nature of the products and the relationship of products with each other (substitutes, complements, etc.)
- More detailed data of customer purchase history. For example, monthly data instead of 5-year aggregates.
- Customer demographics such as gender, age, and income level.

Contents

1 Introduction

1.1 Overall Approach and Basic Assumptions

We were provided a training dataset and a testing dataset in which the variables capture historical customer behaviors prior to the fall of 2008, recorded by the website of a given multichannel retailer. All customers were sent at least one catalog from the website during 2008 fall, and we were given the amount of sales for each customer after they received the catalogs in the training dataset. Our objective is to build a model to predict the amount of sales per customer for this period.

Our overall approach is to create a classification model to predict whether a given customer will make a purchase or not during the promotion period after he/she received the catalogs. We also built a multiple regression model to predict each customer's spending amount, for customers who made a purchase. These models, when run on relevant datasets, will allow us to predict any customer's purchase behavior during the promotion period.

For the purpose of this project, we define outliers as those observations which fall outside 3 standard deviations of all data after accounting for variations due to different sets of predictors. (i.e. a standardized residual of more than 3 in magnitude in multiple regression model). Outlier removal is not necessary for the logistic model as our final model does not involve applicable predictors.

1.2 Report Organization

Our report is structured to follow the general sequence of operations during our modeling process. Chapter 2 illustrates our preparational work for modeling, including data cleaning and exploratory data analysis. During the cleaning phase, we removed observations from data which are either erroneous or irrelevant to our model. To accommodate the classification model, we transformed our response variable `TARGAMNT` into a binary variable. We also performed log-transformations on various right-skewed variables to improve the chances of them making a

better contribution to our multiple regression model. New variables are created from existing ones, including average order amount, purchase rate, breadth of product classes, preferred product class and a frequent buyers indicator. In the exploratory analysis phase, we acquired insights from summary statistics of the variables related to customer purchase behaviors while referring to existing literature for guidelines. Our variables are categorized into those involving recency of activity, frequency of purchase, monetary, product class related, and customer tenure.

Model fitting is the essence of this report. We built our classification model in Chapter 3 to determine whether a customer would make a purchase. We first explored some machine learning techniques in modeling and evaluated each model using common classification accuracy measures. These models provided varying degree of accuracy but none of them served our purpose well. The logistic regression model proved to be the best solution as it provided the best balance in F_1-Score and overall prediction accuracy. For each set of models, we used stepwise regression and best-subset selection to choose the most prominent predictors. Interaction terms were also selectively added to probe for possible ways to increase the predictive power of our model. We concluded that logistic model 3 is the best classification model due to its relatively high performance, small number of predictors and consistent good performance.

We also independently constructed a multiple regression model to predict the sales amount for customers who would make a purchase during the promotion period in 2008. In Chapter 4, we introduced the models we fitted to our training data after further cleaning and variable transformation to best ensure we meet the basic assumptions of linear regression. We produced five different multiple regression models, each basing on certain observations gained in the EDA phase. For each model, we used stepwise regression to systematically removing variables that do not provide additional information in purchase amounts. Observations that appeared to be outliers or with high leverage were removed before the model was re-fitted to ensure that our results are not heavily influenced by a small number of data points. We evaluated each model using adjusted R^2 to assess its goodness of fit. Note that we cannot perform ANOVA tests between models as the input dataset for each model is slightly different due to the outlier

removal process.

Eventually we validated our model against the testing set, analyzed the individual model performance and predicted the top 1,000 customers in Chapter 5. We calculated the expected purchase amount, which takes into account of both the likelihood of purchase and how much the customer is willing to pay, and used that as a criterion for picking the top 1,000 customers.

2 Preparation for Modeling

2.1 Data Cleaning

There are three types of data in the dataset: number of orders, amount of purchase, and time (measured in number of months). Upon inspection, the number of orders and time data do not show any apparent signs of error. For the amount of purchase data, we performed the following data cleaning and transformation:

Invalid Values Some sales data contains negative values and is likely to represent returns. Since we are not modeling customers' behavior after the purchase and negative values represent a small fraction of the dataset, we removed these observations from the dataset.

Transformations All sales data are right-skewed, including our response variable whose distribution is shown in Figure 1 on page 20. We performed log transformation on these variables. After the transformation, the distribution is closer to a normal distribution and the number of observations outside the general spread of the data (3 standard deviations from the mean) is dramatically reduced. See Figure 2.

Outliers To prevent our models from being biased due to the few customers behaving differently from the population, we removed all observations with a standardized residual magnitude over 3 in the multiple regression phase. Outlier removal is not necessary in the classification

phase as our final model involves only categorical variables.

2.2 Exploratory Data Analysis

The variables given in the data set provide us with insights of customers' purchase behavior prior to the Fall of 2008. According to existing research on market analysis[4], customers can be segmented using a few major types of behavioral indicators, some of which are provided in our data. Below we explore the summary statistics of such markers and try to visually determine whether they are correlated with customers' purchase decisions, and the corresponding purchase amount in Fall 2008 using data from the training set.

1. Recency The `recmon` variable measures time since last order and it is measured as of Fall 2008. As seen from Figure 3, the majority of customers (65%) have made at least one order within the last two years. Among the most recent customers (those who made a purchase within the past year), the majority either joined within the past year, or have been with the company for over 3 years (See Figure 4). The most recent customers are most likely to make a purchase and also have a higher average purchase amount in Fall 2008 (see Figures 5 and 6).

2. Frequency We measured frequency using purchase rate, which is calculated as (lifetime orders) ÷ (time on file). Related, we have also defined a binary variable called `frequent.buyer` which takes the value of 1 when a customer's rate of purchase is above 0.1. Frequency gives us insights about how engaged a particular customer is with the products of the company during his/her time with the company. As seen from Figure 7 and Figure 8, likelihood of purchase and corresponding amount of purchase in Fall 2008 increase as frequency increases.

3. Monetary The spending pattern of each customer is likely to be indicative of the amount that the customer would spend in Fall 2008. We examined total monetary value of orders purchased during the customer's lifetime as well as the average monetary value of these orders,

and found them to be positively correlated with the amount that the customer spent in Fall 2008. See Figures 9 and 10.

4. Type of Product Purchased There are a total of 7 product types. Historical data shows that number of purchases and average price vary by product class (see Figures 11 and 12). Figure 13 explores the relationship between whether a customer bought a particular product class in the past and his/her likelihood of purchase in Fall 2008. Class 3 product shows the most distinct pattern: Customers who have bought Class 3 product before are 3 times as likely to buy in Fall 2008 than those who have not bought before. We also looked at the breadth of products that a customer bought in the past (Figure 14). Figure 15 shows that average amount of purchase in Fall 2008 gradually increase as breadth increases and plunges when breadth = 6.

5. Tenure The majority of the customers have been with the company for less than 5 years (Figure 16). We can see from Figures 17 and 18 that customers who joined recently and customers who have been with the company for a long time tend to have a higher likelihood of purchase. The same trend appears when we consider the average purchase amount instead of percent of customers making a purchase.

6. Past Promotion Purchases Similar promotions like the one in Fall 2008 have been held in previous years. Figure 19 shows that customers who have purchased in earlier fall promotions are more likely to buy again this year. The relationship is the strongest with the 2007 promotion.

Summary Table 1 lists all variables we considered which were included in our model, grouped by aspects of purchasing behavior as discussed above. Both the variables provided in the datasets and the variables we created from the existing ones are included.

3 Classification Models

The first part of our project requires a classification model. This is because over 95% of our customers did not make a purchase during the most recent promotion period. Including these customers in a model that predict sales amounts will lead to the model being heavily biased towards predicting low or no purchase at all. Therefore, we first fit a classification model that allows us to identify customers who will make a purchase.

We will be comparing the different models through classification accuracy measures, the most comprehensive of which is the F_1 score[3]. All models will use the same binary response variable indicating whether a sale is made. For the purpose of evaluating models, we broke the "training" set into a fitting set and a validation set. All models are fitted on the fitting set and then evaluated on the validation set, but the final model is fitted on the entire training set before making predictions. In order to rule out models that do not give us any information, we first created a baseline case using a random classifier, such that each observation is randomly assigned a label with the proportions of sales and no-sales as its probability. As we can see from appendix table 2, just randomly assigning labels gives us a 89.5% accuracy because of the highly unbalanced number of observations with each label.

3.1 Machine Learning Models

One of the simplest machine learning algorithm, Naive Bayes, has been successful in classification in many different fields. As different algorithms are used depending on the form of input, it is hard to mix binary, multilevel categorical, and continuous predictors. We chose to use the five indicator variables that were most highly correlated to making a sale: `ord185`, `ord285`, `ord385`, `ord485`, and `freq.buyer`. This model gave us a higher precision and accuracy in general, but due to the heavily unbalanced classes a very low recall and in turn an unsatisfactory F_1 score.

We proceeded using support vector machines (SVM) as our next candidate. SVM is a much

more sophisticated modeling technique than Naive Bayes, and is in general very good at separating classes even when the decision boundary is highly non-linear. We again trained the model with all predictors as well as the 5 core predictors, but SVM performed poorly again due to the unbalanced classes, giving a good precision but inadequate recall.

A third approach we attempted is using random forest, which is an ensemble learning method that involves randomized decision trees. Because this algorithm involves random sampling, we are able to manually adjust for the unbalanced classes, such that during each iteration of the learning process a balanced subset of the training data is used. However, because the larger class (no sales) has much bigger variability in most predictors, this approach in fact produced very low precision with much higher recall. This is even less desirable than the previous cases, as it lowers our prediction accuracy to even worse than a random classifier.

3.2 Logistic Regression Model

As all of our machine learning models provided unsatisfactory results, we continued to explore our data using logistic regression. An additional advantage of using logistic regression is that we are able to estimate the probability of purchase for each customer, which will enable us to probabilistically determine the expected revenue for the entire promotion period, rather than by selecting customers first (as is the case for models above).

For the purpose of evaluating each model, we also calculated classification measures for all models. As our data is heavily biased towards no-sales, we assign labels using a threshold probability of 0.2. In other words, we only predict a sale if the customer has a probability of making purchase greater than 0.8. This number is chosen as it maximizes the F_1 measure and is best for comparing the models. This will not directly affect our final prediction, as we will use the probabilities directly instead of thresholding it into a binary prediction. As an additional measure of goodness of fit, we also compute the McFadden pseudo-R^2 for each model[5].

1. Full Model Our first logistic model is simply the full model, using all applicable predictors. This gives us a sense of the upper bound of our predictive power. This model gave us pseudo-R^2 of only 14.3%, which basically disallows any better predictions unless we can find better transformations and new variables. More diagnostic statistics for this model are listed under model 1 in appendix table 3, and the output for this model is in appendix output 1.

Since most of the predictors follow extremely skewed distributions, we also attempted using $\log$ transformations to normalize them. However, none of the transformed variable resulted in an increase in pR^2 of more than 0.01. Considering that the logistic model would be much harder to interpret when these $\log$-transformed variables are used, we decided to use all variables as-is.

2. Variable Selection through Stepwise Regression We can see from output 1 that most of the predictors are not statistically significant. This is expected as many of them are highly correlated to each other. To eliminate extraneous predictors, we performed stepwise regression using AIC to generate a smaller model. This is model 2 listed in table 3. Predictors retained in this model include total lifetime sales, 5-year sales in classes 1, 2, 3, 5 and 6, 5-year orders in classes 2, 3, 4, and 5, time since last order, and the frequent buyer indicator. Regression output for this model is shown in output 2. Although the stepwise algorithm concludes this would be the best combination of predictors, it is important to note that sales of class 1 as well as total lifetime sales are not significant.

We also tried best subset selection. The predictors included in this model are all `ordcls` and `ord85` variables with `recmon`. Although this method gave us similar results as stepwise regression, it took much longer time as the algorithm exhaustively fitted every possible models. So we decided not to proceed with this method any longer.

3. Further Reduced Model As we would want to maximize our prediction results rather than pR^2, we can continue to remove variables as long as our prediction accuracy is not impacted. After iteratively removing the least significant features, we arrive at a highly simplified model with only 5 predictors, all of which are binary: Purchase during the 85 promotion period within

the past 4 years, and whether the customer is a frequent buyer. Surprisingly this model (number 3 in table 3) achieved marginally better prediction than all other models, while still maintaining comparable pR^2. This is definitely a much better model than the one obtained through stepwise regression, as very little information is lost while removing a large group of predictors. Regression output for this model can be found in output 3 on page 32.

4. Models with Interactions We have acquired a model with acceptable predictive power. A good way to possibly improve it is through the use of interaction terms. Since we know that `ordcls3` is highly correlated with many other of our attributes, we chose to use it to interact with other variables. We first created another logistic model basing on model 3 above and adding all interaction terms involving `ordcls3`. Prediction statistics are shown in table 3 under model 4. We then performed stepwise feature selection on this model to pick the best interactions, which is our model 5. From output 4, we can see that this model in essence brought back all predictors from model 2 that were dropped. Although they are all statistically significant, they again contribute marginally to our classification, providing only 0.1% in additional raw accuracy and 0.007 in F_1 measure

3.3 Verdict and Reflection

Comparing all of the above models, we believe that a logistic model will be the best in estimating expected revenue. We conclude that Model 3 is the best as it involves the least number of predictors while still achieving prediction rates that are comparable to the other models.

This model makes intuitive sense, as we have seen during our EDA phase that customer purchase behavior during each promotion period is highly correlated to their purchases during previous sales periods (see Figure 19). Therefore, it is reasonable that they end up being the most important predictors of sales.

One potential downfall of this model is that as the variability in our data is very large and uneven, the performance of each model depends heavily on the training dataset used. All model

diagnostic measures presented in this section are computed using the same training and validation sets and are comparable to each other. After inspecting many randomly drawn samples, we found that the diagnostic measures vary wildly, but the relationships between models are similar. The model we have chosen consistently perform well, while some of the larger models seem to suffer from overfitting with prediction rates fluctuating wildly. Therefore, our chosen logistic model provides a relatively accurate and reliable estimation of customers' purchase behavior.

4 Multiple Regression Models

After having a model to predict whether a customer would make a purchase in the most recent promotion, our next step is to predict the sales amount for those customers who would make a purchase. This will enable us to capture another aspect of customer's value, *i.e.* the amount of money that he/she is likely to spend on the purchase.

4.1 General Approach

Because the response variable is a continuous variable, we used a multiple regression model. We first subset the training data to include only observations where `targamnt` > 0. 2,825 out of the original 52,844 observations satisfy this criterion. We then removed negative values and performed log transformation on the sales-related variables as described in Section 2.1. This leaves us with 2,822 observations in the training set.

When selecting predictors for the model, we considered all of the variables listed in Table 1. For each model, we started with some initial set of predictors which we think are good candidates for predicting sales amount. We did not include predictors that are likely to correlate with each other because that would cause our model to be prone to the problem of multicollinearity. For example, we did not include the average order amount by class variables and the sales amount by class variables at the same time because the average order amount by class variables are derived from sales amount by class variables and they are likely to be strongly correlated

with each other.

After we chose the initial set of predictors, we ran them through stepwise regression using AIC to eliminate extraneous predictors. We then performed diagnostics on the model to check whether it satisfied model assumptions, which include homoscedasticity, normality, and absence of multicollinearity, outliers, and influential observations.

4.2 Model Comparison

Model 1: Model with Order and AOA Variables In the first model, we started with order variables (`ordcls1-ordcls7`, `totord`, average order amount variables (`AOA`, `AOA1 - AOA7`), past year promotion variables (ord185-ord485), total sale, time on file, and purchase rate. The full model gives an R^2_{adj} of 0.1876. Summary result shows that some predictors are insignificant, suggesting that the model can be further reduced. After performing stepwise regression using AIC, 16 of the original 24 variables were retained. The reduced model gives an R^2_{adj} of 0.1885.

There are 11 outliers (*i.e.* observations whose standardized residuals are above 3 in absolute value). We removed the outliers from the training data and repeated the fitting steps above until all outliers were removed. The model built on data with no outliers gives us an R^2_{adj} of 0.2043. All predictors are significant at the 5% significance level except for ln(`AOA5`) and `PR`, which are significant at the 10% significance level.

The variance inflation factors (VIF) of all predictors in Model 1 are below the threshold of 10, suggesting that multicollinearity is not present. Cook's Distance for all observations are below 1, which shows that no observations are influential. To test for homoscedasticity, we plotted the residuals against the fitted values. The plot (see Figure 20) is roughly a horizontal band around 0, which confirms that the homoscedasticity assumption is satisfied. For the test on normality, we plotted the standardized residuals against theoretical normal distribution quantiles (Figure 21), which shows a roughly straight line, suggesting that the normality assumption is satisfied.

Model 2: Model with Sales and Breadth Variables In the second model, we started with the same set of variables in Model 1 but replaced the `order` and `AOA` variables with the sales variables (`salcls1 - salcls7`) and the `breadth` variable. After running stepwise regression, removing outliers, and confirming that model assumptions are satisfied, the final Model 2 has 12 predictors and it gives an R^2_{adj} of 0.1669. Most predictors are significant at the 5% significance level. Starting with the sales variable instead of the AOA and order variables gives us a more succinct model. However, with an almost 20% reduction in R^2_{adj}, the decrease in predictive power outweighs the benefit of having a more parsimonious model. We therefore conclude that Model 2 does not outperform Model 1.

Model 3. Model with Sales and Prefer Variables In the third model, we started with the same set of variables as those in Model 2 but replaced the `breadth` variable with another representation of product class, the `prefer` variables. Following the same procedures as described above for Model 1 and Model 2, the final Model 3 has 11 predictors and it gives an R^2_{adj} of 0.1523. Only the `prefer1` variable (Class 1 Product being the preferred product) is retained in this model. It is worth noting that `ord285` and `ord385` are not significant at even the 10% significance level but are nevertheless retained in the model. Like the previous model, our current model has lower predictive power after adjusting for the number of predictors, so Model 1 remains the best model.

Model 4 and 5. Models with Interactions One way to possibly increase the predictive power of the model is through the use of interaction terms. We started with variables in Model 1 and added interactions terms that are likely to be significant from our understanding of the data. We first tried adding the interaction of time on file variable with all the `AOA` variables. Our model ended up with 17 predictors and an R^2_{adj} of 0.2046. The only retained interaction term, ln(`AOA6`):`tof`, is not statistically significant. Compared with Model 1, the increase in R^2_{adj} is negligible. Model 1, being the simpler model, therefore remains our top choice.

We also tried adding in the interaction of purchase rate with all the `AOA` variables. Because

the presence of both PR interaction terms and PR variable causes multicollinearity, we removed PR variable from the model while keeping the PR interaction terms. Again after our standard modeling procedure, this model could explain 20.67% of the variation in target-window sales. The interaction term ln(AOA3):PR, kept by stepwise regression, is highly significant. Again, this model did not provide us with a substantial increase in predictive power, and we chose not to use this as our final model.

4.3 Verdict and Reflection

By comparing the performance of the above models and their model diagnostics results (See Table 4 and Output 5 through Output 9 in the Appendix), we conclude that Model 1 is the best because it has the highest predictive power, its number of predictors are not much greater than those of other models, and it satisfies all the multiple regression assumptions.

We also performed 10-fold cross-validation for each of the above models to validate our regression results and to ensure randomness in our training sample selection. The outcome was as expected, with models 1,4 and 5 having the least Mean Square Error across all folds. (See Table 5 in Appendix)

As seen from the regression result output of Model 1, variables representing the monetary aspects of a customer's purchasing behavior are highly significant (1% significance). For example, target-window sales is strongly and positively correlated with average order amount(AOA) and lifetime total purchases(totsale). An increase in tenure (tof) and frequency (PR) also lead to an increase in target-window sales on average, though the significance level is slightly lower (5% and 10%). In contrast to the logistic model, orders in past promotions are negatively correlated with sales, suggesting that customers who bought in the previous promotion on average spent less in this year's promotion.

The effect of product class on sales is mixed. Average order amount of certain product classes (*e.g.* AOA3, AOA5, and AOA7) are positively correlated with sales in the recent promotion, while others (*e.g.* AOA2 and AOA6) exhibit a negative relationship with target-window sales. For

some products, the past average order amount (AOA) plays a bigger role in determining target-window sales compared to past orders (`ordcls`). For example, for Product Class 1, past order is a significant predictor, but past AOA is not. Information about these different types of products and their relationship with each other (*e.g.* whether they are substitutes or complements) can provide us with insights to better understand the regression output and also to build a more rigorous model.

5 Prediction

5.1 Model Validation Against Test Set

In order to estimate sales for customers in our testing set, we used the logistic model to predict the probability of the customer making a purchase, and used the multiple regression model to estimate the sales amount should the customer make the purchase. We then multiplied the two together to obtain the expected purchase amount for each customer. This is because

$$E(\text{Amount}) = P(\text{sale}) \cdot E(\text{Amount}|\text{Sale} = 1)$$

Expected purchase amount takes into account two aspects of a valuable customer: his/her likelihood of purchase and how much the customer would spend should he/she make the purchase. Since our multiple regression model is only fitted on the set of customers with positive sales, it is likely to be upward biased. Multiplying the predicted sales amount by probability will help adjust the predicted sales downwards.

After estimating each customer's spending, we used different error measures to evaluate our predictions on the testing set. Our model provided estimations that yielded the following values:

$$\begin{aligned}
\text{Sum of Squared Errors (SSE)} &= \sum_i \left[E(\text{Amount}_i) - \text{Amount}_i\right]^2 && = 50,475,637 \\
\text{Mean Squared Errors (MSE)} &= \text{SSE}/n && = 946.8 \\
\text{Root Mean Squared Errors (RMSE)} &= \sqrt{MSE} && = 30.8 \\
\text{Mean Signed Deviation (MSD)} &= \frac{1}{n}\sum_i \left[E(\text{Amount}_i) - \text{Amount}_i\right] && = -0.96 \\
\text{Mean Absolute Error (MAE)} &= \frac{1}{n}\sum_i \left|E(\text{Amount}_i) - \text{Amount}_i\right| && = 7.30
\end{aligned}$$

We regard these results to be satisfactory. The RMSE value shows that accounting for variables in our prediction models, the standard error of our prediction on customer spending during promotion is $30.8. In other words, our model is able to estimate spending amount to within $30 for at least half of our customers in the testing set. On average, this model makes a prediction error of $7.30, and it tends to predict an amount slightly lower than actual as suggested by the MSD.

5.2 Individual Model Performance Analysis

The logistic regression model correctly classified 93.5% of the customers in the testing set as making a purchase during the Fall promotion period. Our model achieved a precision of 0.342, recall of 0.237, and an F_1 score of 0.280. These measures are all within the range of variations we saw while cross-validating our models, which shows that the testing data behaves similarly to our training data. Additionally, this also means that our model is predicting customer behavior similarly in the testing set, and thus our model does not over-fit particular datasets.

We computed individual performance of the multiple regression model by computing the prediction error for customers who actually made a purchase in the most recent promotion.

By focusing on customers who have positive actual sales, we filtered out errors introduced by incorrect classification. Multiple regression model alone has the following error measures:

$$\text{SSE} = 45,547,722 \qquad \text{MSE} = 15,903.53$$

$$\text{RMSE} = 126.11 \qquad \text{MSD} = -14.30$$

$$\text{MAE} = 45.98$$

5.3 Predicting the Top 1,000 Customers

Our next task is to select 1,000 customers who are likely to spend the most in the recent promotion. We picked the customers with the highest expected purchase amount based on our prediction as our top 1,000 customers. The total predicted purchase amount for these 1,000 customers is $174,238.4. The total actual purchase amounts by the true top 1,000 customers is $165,965, which is relatively close. 443 out of the 1,000 predicted top customers eventually made a purchase in the target window. They spent $ 57,432.14, which amounts to 34.6% of the total amount spent by the true top 1,000 buyers.

6 Conclusion

Through exploratory data analysis and model fitting, we find that frequency of purchase and purchases during prior year promotions are important predictors for whether a customer makes a purchase or not in the recent promotion. This makes sense because we would expect a customer who frequently buys the company's product to make a purchase during this promotion period as well. Similarly, those who made a purchase in past promotions are more likely to show interests in this year's promotion.

The significant variables in our best multiple regression model are quite different from those

in the logistic regression model. This is probably due to the fact that predicting the actual sales amount and predicting whether a customer would make a purchase depend on different aspects of customers' purchasing behavior. Variables representing monetary aspects of a customer's purchasing behavior such as the average order amount(`AOA`) and lifetime total purchase amount(`totsale`) are highly significant in the multiple regression model. This implies that customers who spent more per order in their past purchases are highly likely to also spend more during this coming promotion period.

There are some variables which we think would be good predictors to improve our models. First, it would be better if we could have more background information about the types of the products that the company is selling. As observed in the verdict and reflection part in Chapter 4, some product classes are positively correlated with the sales amount while others are negatively correlated. Information about the product classes will enable us to appropriately group certain product classes together and in turn build a more robust model.

Second, the data that we have for historical sales and orders are on the aggregate level. If more granular data (say monthly data) are available to us, we can enhance our predictions by incorporating information about sales trends such as growth over time and seasonality patterns into our models. Having more detailed data will also enable us to ascertain whether a particular purchase happens during a promotion or a regular sales period and control for the effect of promotion on sales.

Last but not least, customer demographics such as gender, age, education level, and household income, would be useful to help us segment the customers, make better predictions, and most importantly, identify and target the most valuable customers.

Appendix A Tables

Table 1. List of Variables Considered

Category	Variables
Recency	`recmon`(Time since last order)
Frequency	`PR`(Purchase Rate) `freq.buyer`(Frequent Buyer)
Monetary	`totsale` `salcls1` - `salcls7` `AOA`, `AOA1` - `AOA7` [a]
Product Type	`ordcls1` - `ordcls7` `prefer1` - `prefer7` [b] breadth [c]
Tenure	`tof`
Past Promotion	`ord185` - `ord485`

[a] Average Order Amount (AOA): (sales amount) ÷ (number of orders) overall or in a product class. When no orders were made in a class, we set AOA = 0.

[b] Preferred class: binary variable indicating whether a product class is the one which the customer made the most number of orders among all purchases.

[c] Breadth is defined as the total number of product classes that a customer has bought in the past.

Table 2. Comparison of Machine Learning Models
Shaded cells indicate best model by each criteria

Model	p [a]	Precision	Recall	F_1	Accuracy
Baseline - Random Classifier	0	0.064	0.057	0.061	0.895
Naive Bayes	5	0.480	0.177	0.259	0.939
SVM 1	24	0.611	0.083	0.147	0.942
SVM 2	5	0.614	0.113	0.192	0.943
Random Forest 1	24	0.156	0.648	0.252	0.771
Random Forest 2	6	0.159	0.654	0.257	0.775

[a] Number of predictors in models

Table 3. Comparison of Logistic Models
Shaded row indicates model chosen

Model	p	Precision	Recall	F_1	Accuracy	Pseudo R^2
1	24	0.357	0.361	0.359	0.923	0.143
2	15	0.357	0.360	0.358	0.923	0.142
3	5	0.359	0.355	0.357	0.924	0.135
4	28	0.357	0.360	0.358	0.923	0.140
5	12	0.365	0.360	0.364	0.925	0.139

Table 4. Comparison of Multiple Regression Models
Shaded row indicates model chosen

Model	p	Multiple R^2	Adjusted R^2
1	16	0.2088	0.2043
2	12	0.1704	0.1669
3	11	0.1556	0.1523
4	17	0.2095	0.2046
5	16	0.2112	0.2067

Table 5. 10-Fold Cross-Validation Comparison of Multiple Regression Models
Shaded row indicates model chosen

Model	MSE
1	0.361
2	0.376
3	0.385
4	0.361
5	0.36

Appendix B Figures

Figure 1. Distribution of Sales Amounts, Fall 2008

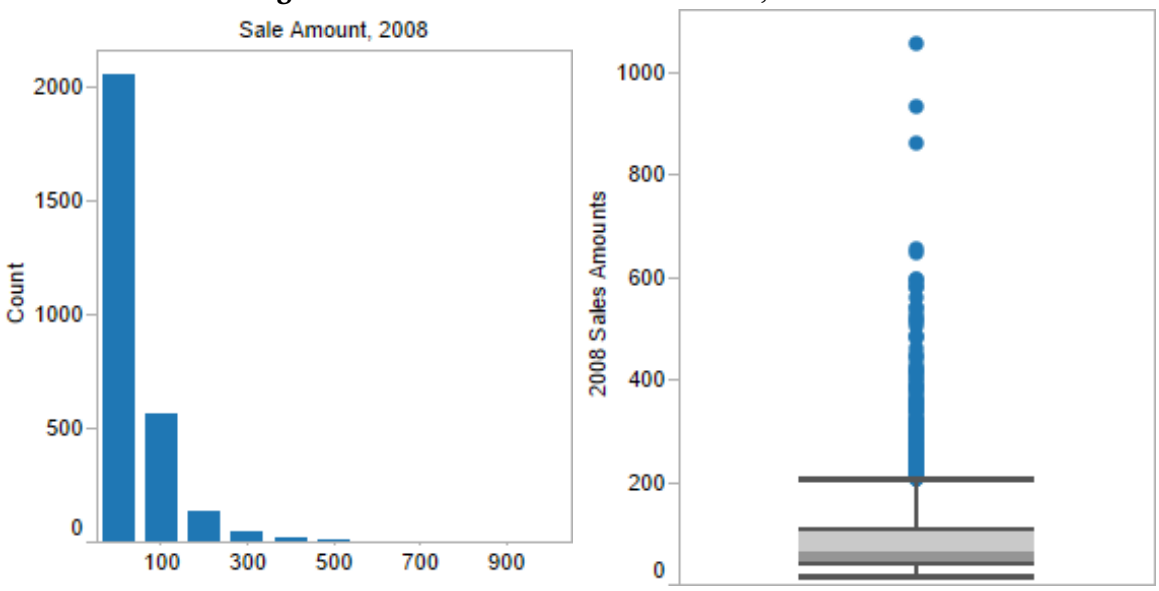


Figure 2. Distribution of Log-Transformed Sales Amounts, Fall 2008

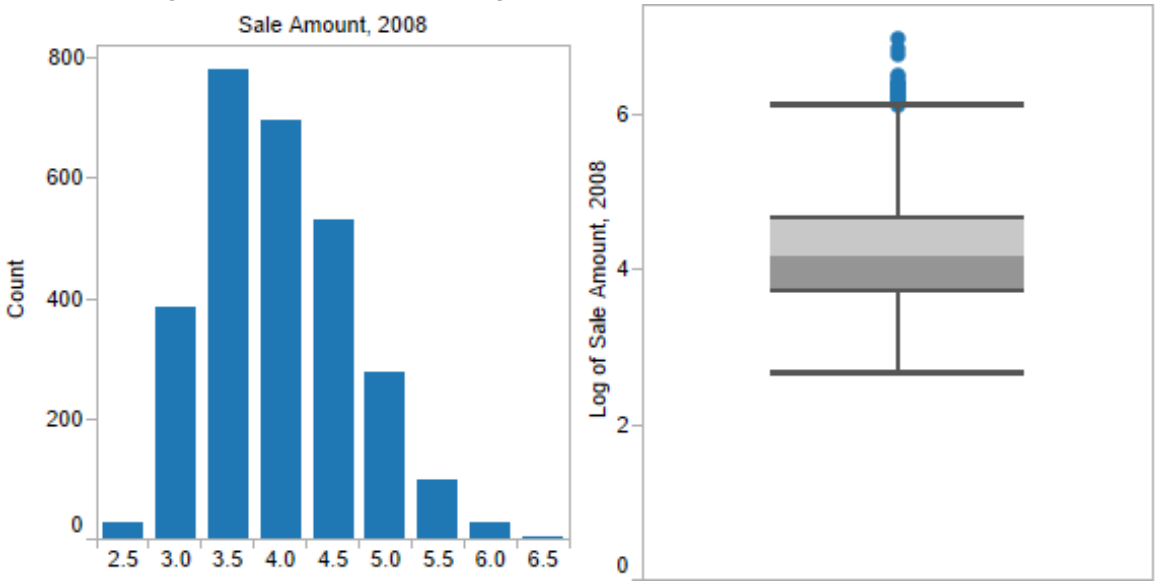


Figure 3. Distribution of Customers by Years since Last Order

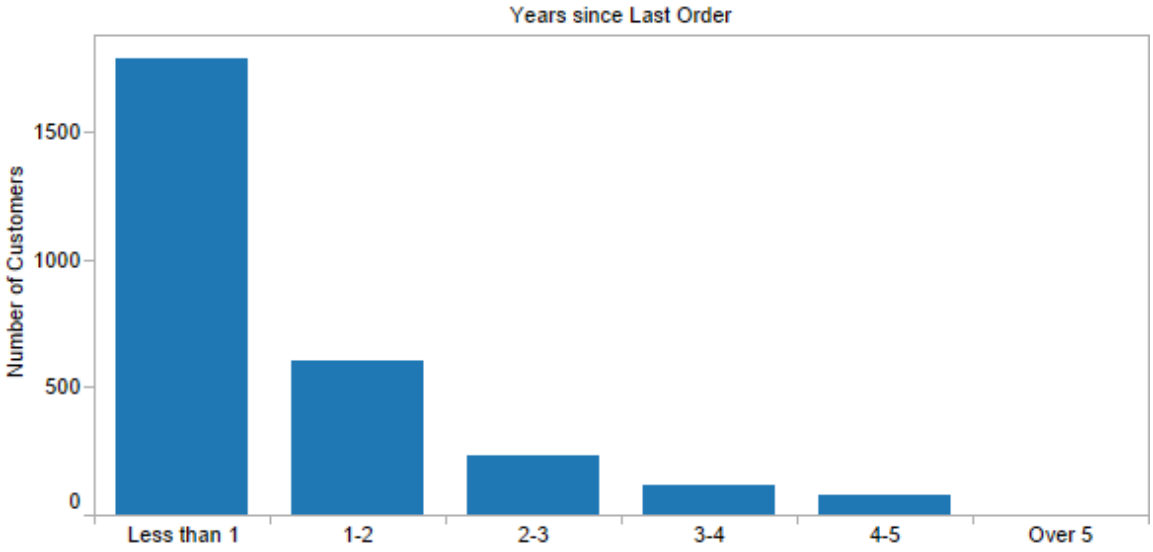


Figure 4. Distribution of Customers whose last purchase was within a year by Year on File

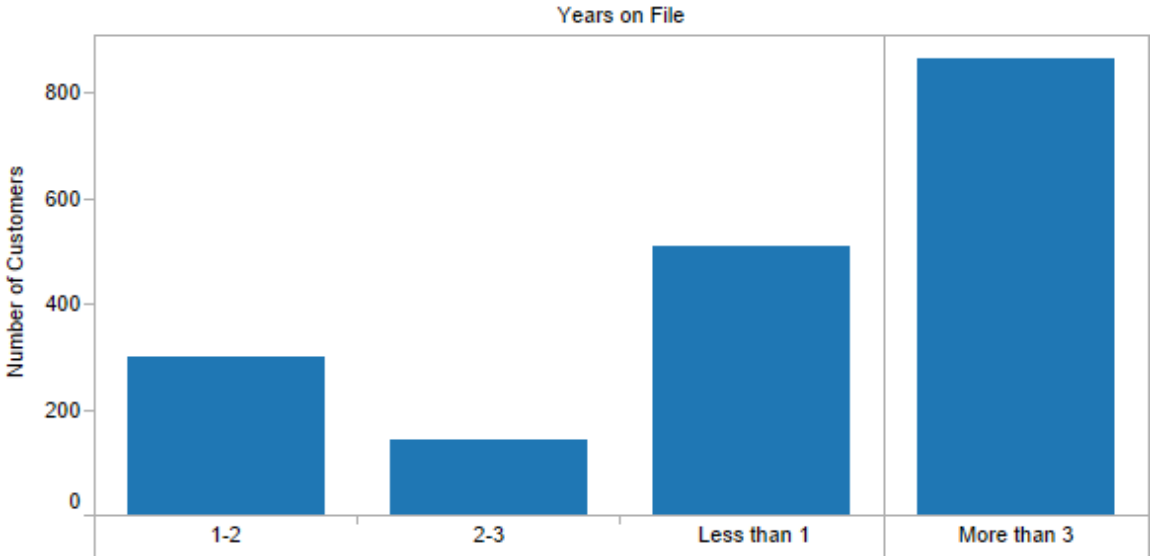


Figure 5. Percent of Customers Making Purchase by Years Since Last Order

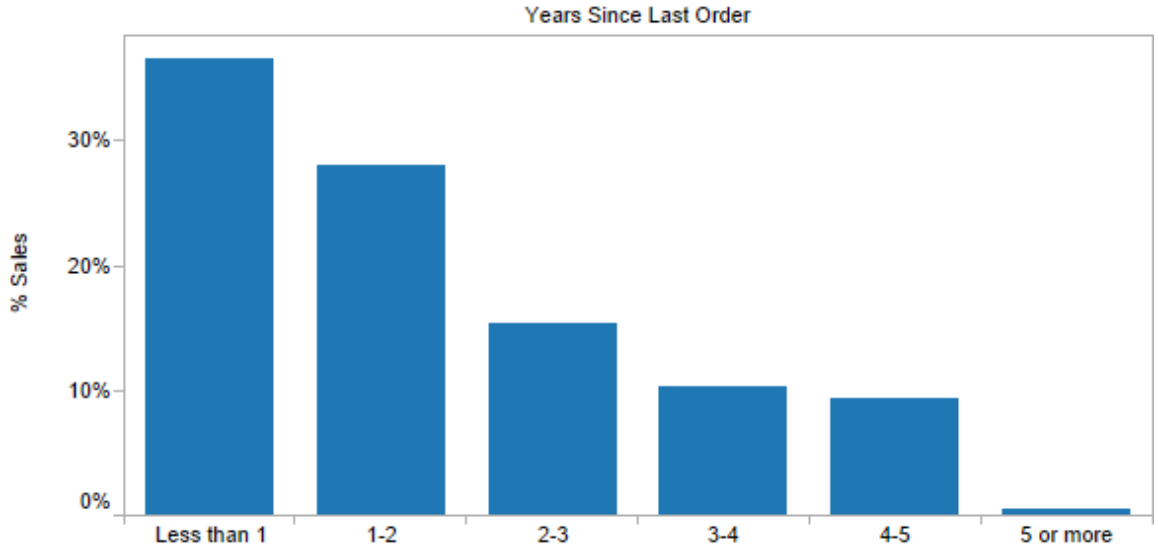


Figure 6. Sales Amount by Customers' Years Since Last Order

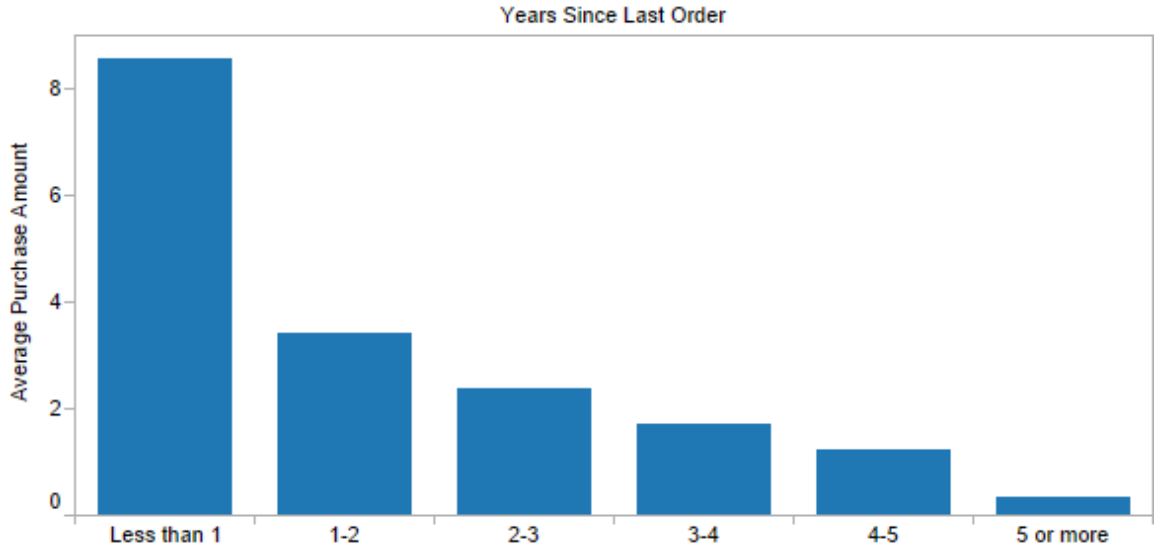


Figure 7. Percent of Customers Making Purchase by Historical Purchase Rate

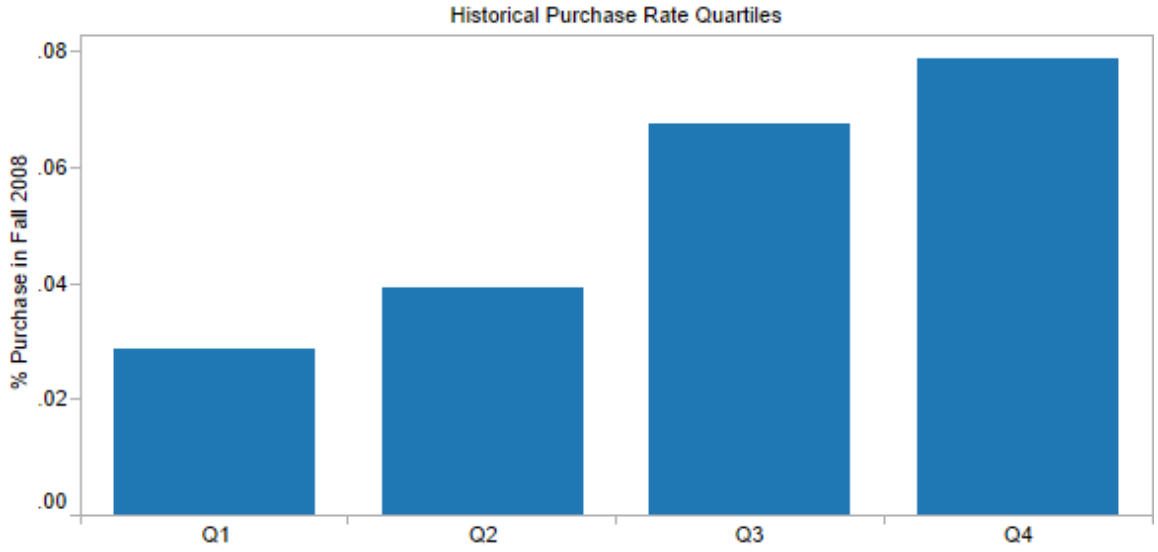


Figure 8. Sales Amount by Customers' Historical Purchase Rate

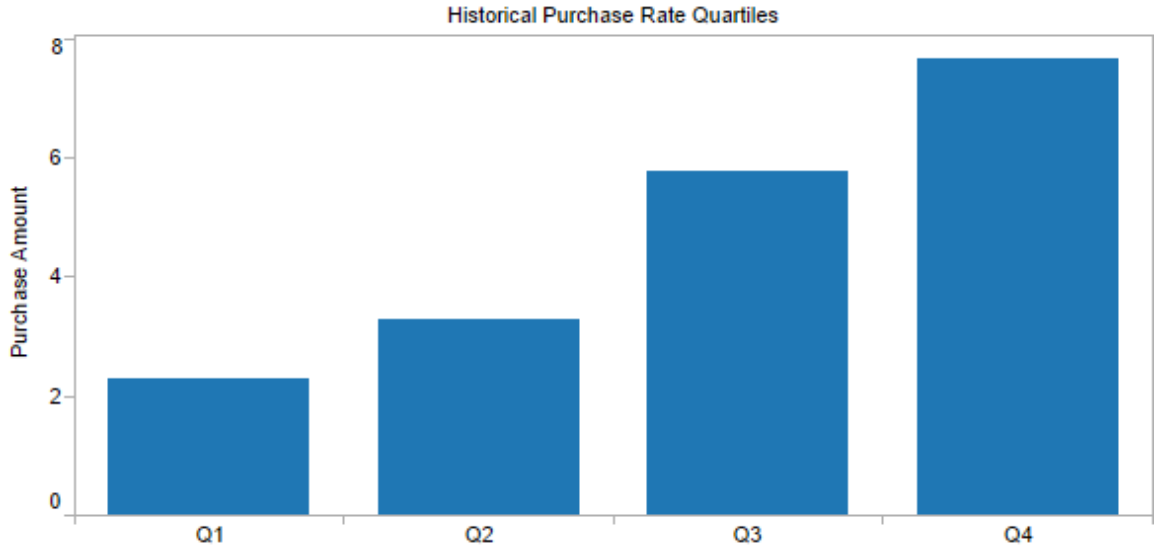


Figure 9. Average Order Amount vs. Sales in Target Window
(Log-Log Scale)

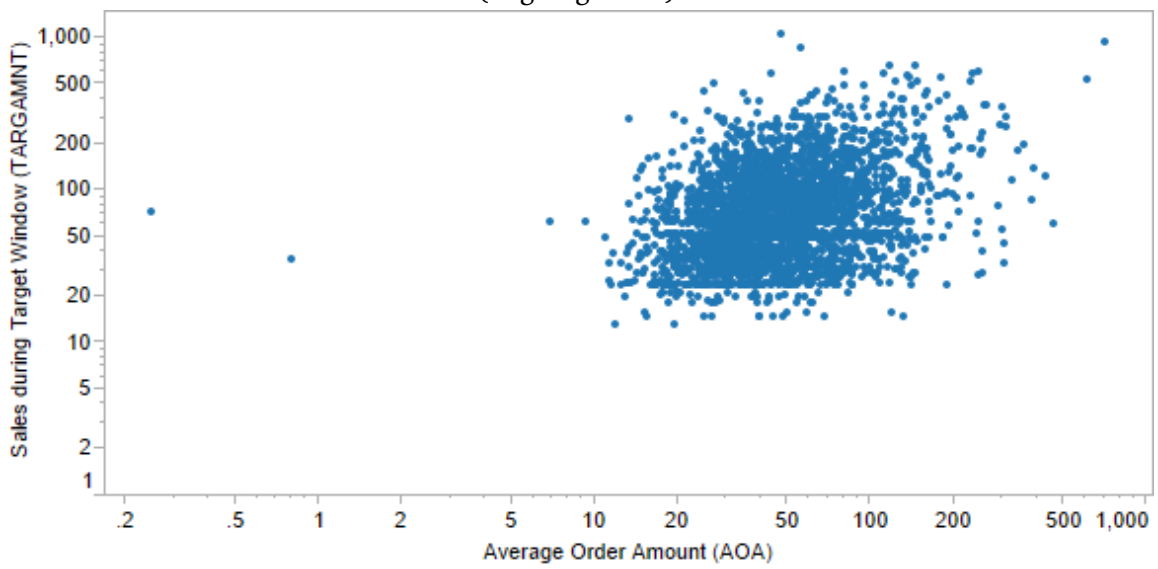


Figure 10. Total Sales vs. Sales in Target Window
(Log-Log Scale)

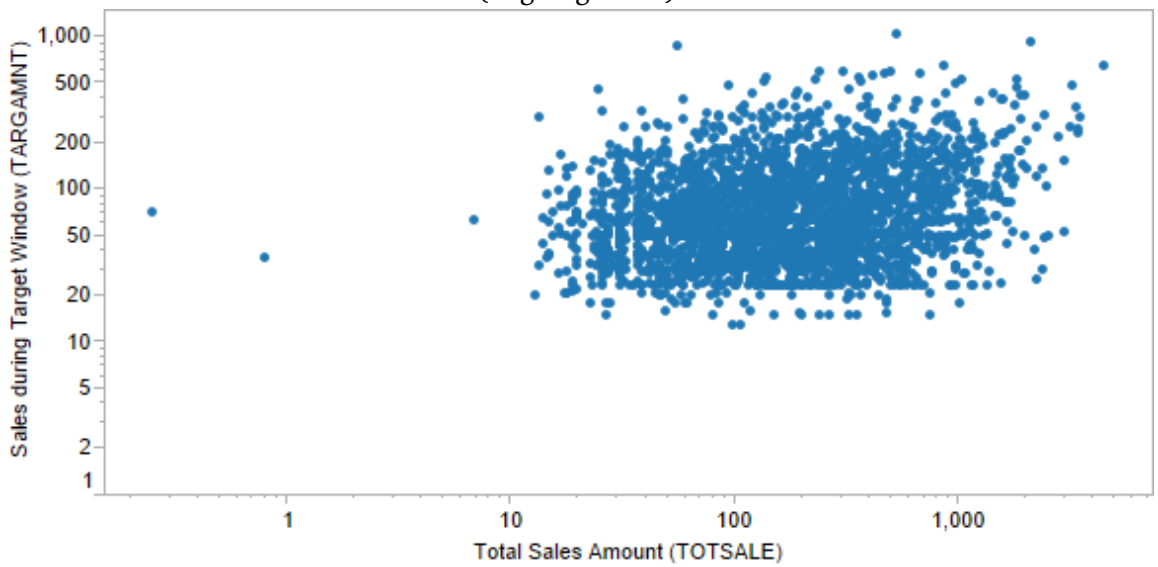


Figure 11. Number of Orders by Product Class

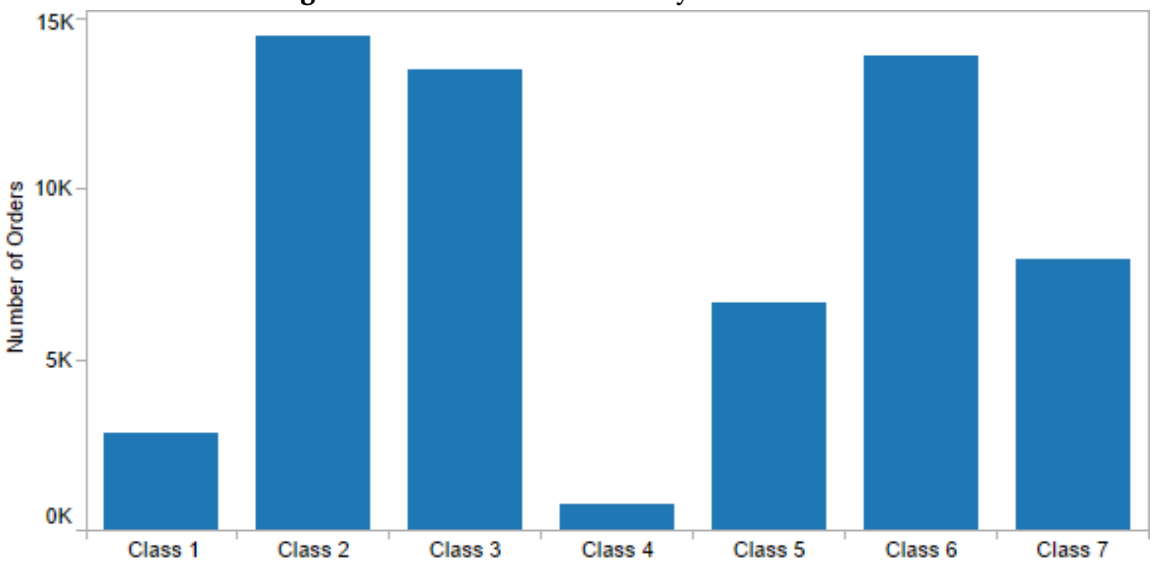


Figure 12. Item Price by Product Class

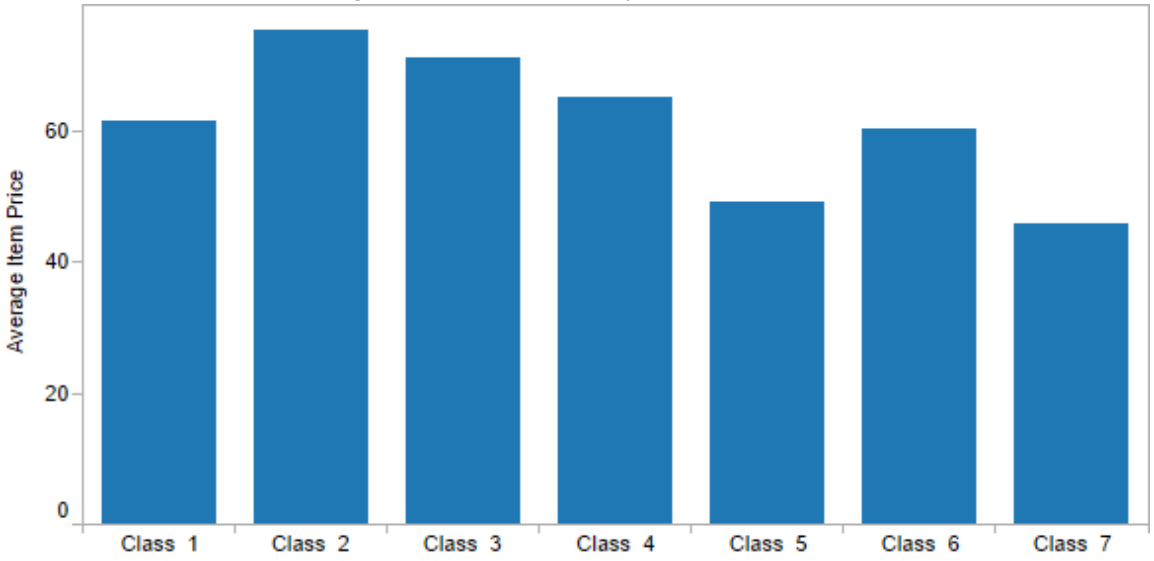


Figure 13. Percent of Customers Making Purchase by 5-Year Purchase Behavior in each Product Class

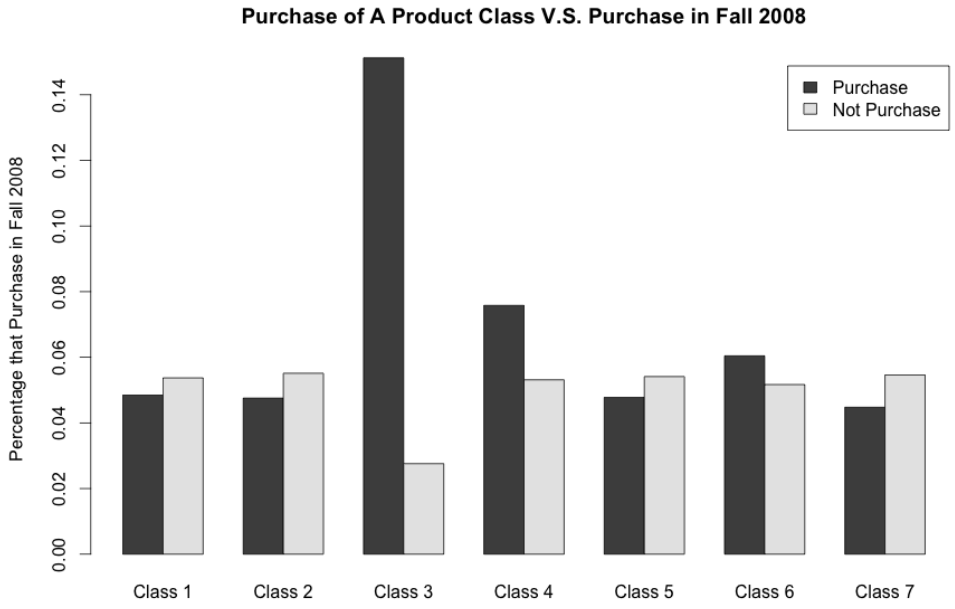


Figure 14. Number of Orders by Customers' Historical Breadth of Purchase

Figure 15. Average Purchase Amount by Customers' Historical Breadth of Purchase

Figure 16. Distribution of Customers by Time on File

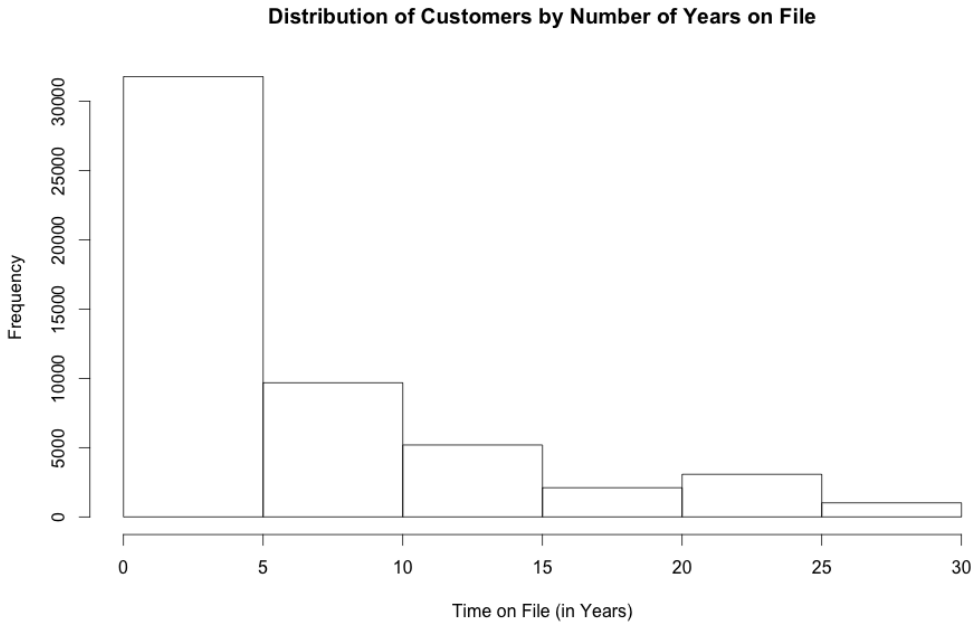


Figure 17

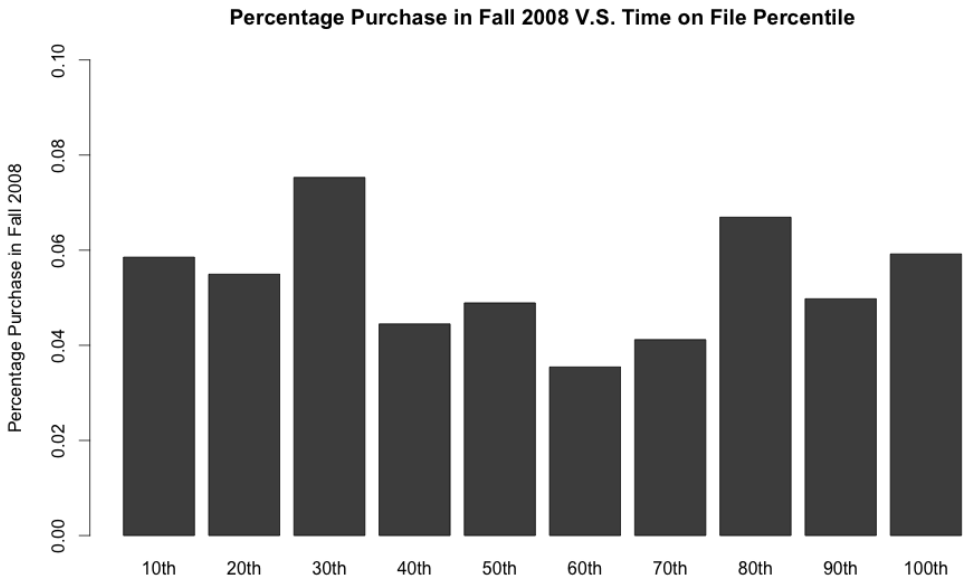


Figure 18

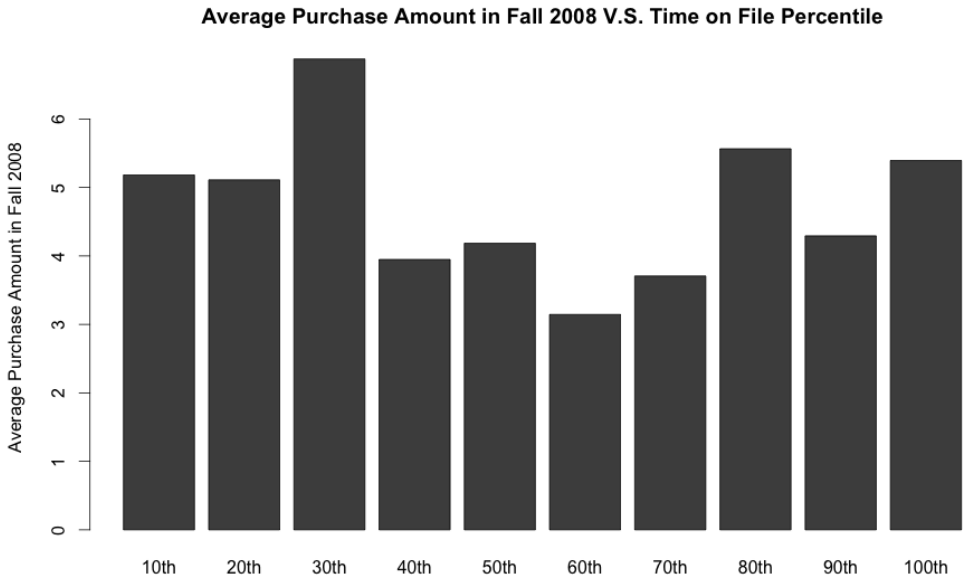


Figure 19

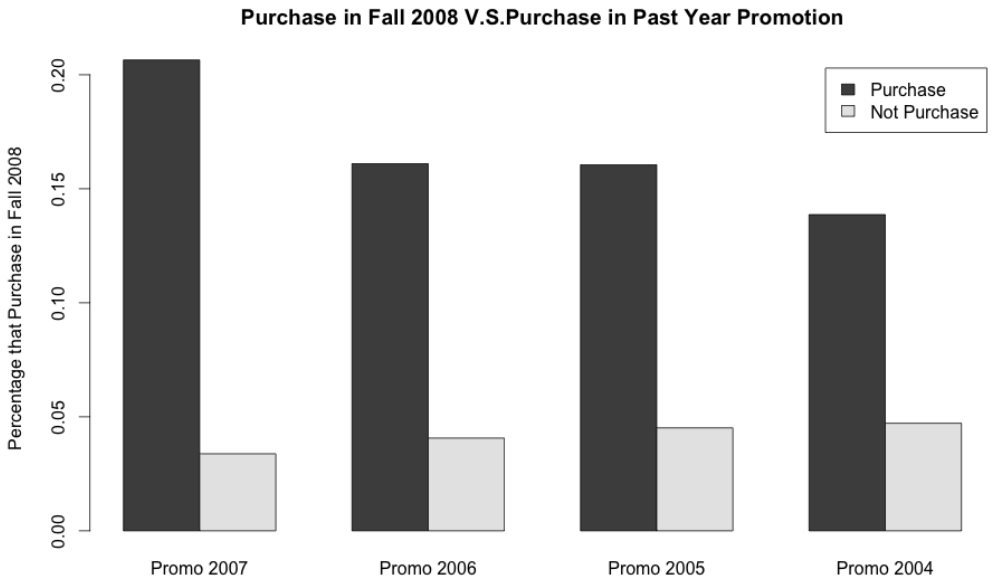


Figure 20. Residuals vs Fitted Values Plot for Multiple Regression Model

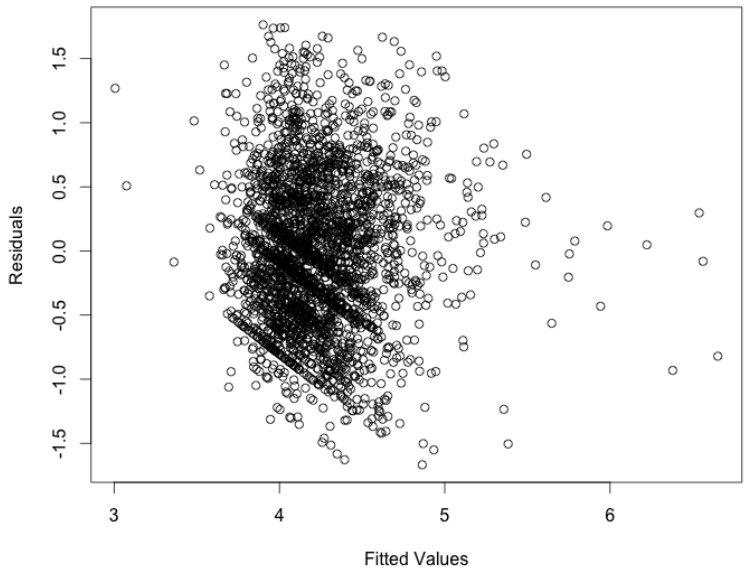
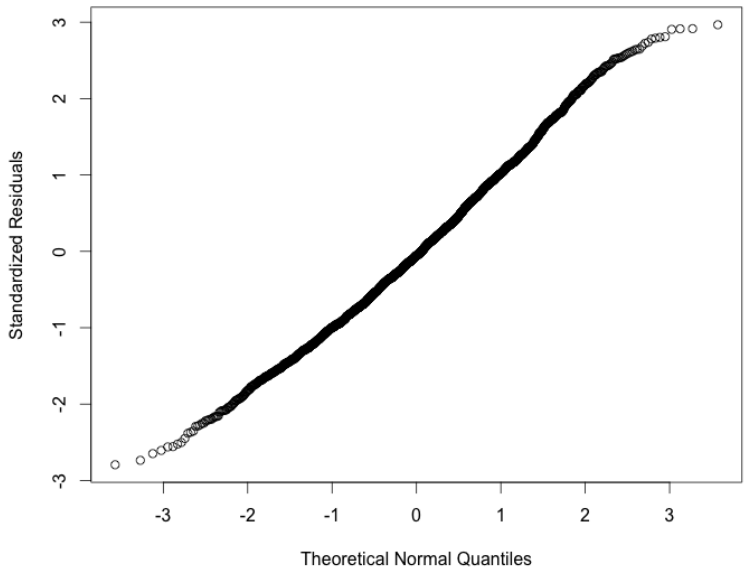


Figure 21. Standardized Residuals vs Theoretical Normal Quantiles for Multiple Regression Model



Appendix C R Output

Output 1. Full Logistic Model

```
Call:
glm(formula = sale ~ recmon + tof + ordcls1 + ordcls2 + ordcls3 +
    ordcls4 + ordcls5 + ordcls6 + ordcls7 + salcls1 + salcls2 +
    salcls3 + salcls4 + salcls5 + salcls6 + salcls7 + ord185 +
    ord285 + ord385 + ord485 + totord + totsale + avgordamt +
    freq.buyer, family = binomial, data = traintrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3167  -0.2899  -0.2385  -0.2013   2.9442

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.4638308  0.0841575 -41.159  < 2e-16 ***
recmon         -0.0146955  0.0026740  -5.496 3.89e-08 ***
tof             0.0004176  0.0005191   0.805  0.42110
ordcls1        -0.0852285  0.1776687  -0.480  0.63144
ordcls2         0.0590113  0.0593490   0.994  0.32007
ordcls3         0.2314039  0.0825607   2.803  0.00507 **
ordcls4         0.3256444  0.2120296   1.536  0.12458
ordcls5        -0.1830928  0.0923731  -1.982  0.04747 *
ordcls6         0.0463937  0.0572922   0.810  0.41807
ordcls7        -0.0868465  0.0652557  -1.331  0.18323
salcls1        -0.0013363  0.0023704  -0.564  0.57292
salcls2         0.0004974  0.0005634   0.883  0.37729
salcls3         0.0022494  0.0005224   4.306 1.66e-05 ***
salcls4         0.0005235  0.0017098   0.306  0.75948
salcls5         0.0026276  0.0012522   2.098  0.03588 *
salcls6         0.0013584  0.0006797   1.999  0.04564 *
salcls7         0.0003416  0.0006156   0.555  0.57890
ord1851         1.1758268  0.1040427  11.301  < 2e-16 ***
ord2851         0.6107320  0.1053505   5.797 6.75e-09 ***
ord3851         0.8311219  0.0712872  11.659  < 2e-16 ***
ord4851         0.7662486  0.0899170   8.522  < 2e-16 ***
totord          0.0037499  0.0077748   0.482  0.62958
totsale        -0.0003074  0.0001923  -1.598  0.10994
avgordamt       0.0005876  0.0003958   1.485  0.13765
freq.buyerTRUE  0.1654021  0.0682303   2.424  0.01534 *
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16273  on 39999  degrees of freedom
Residual deviance: 13950  on 39975  degrees of freedom
AIC: 14000

Number of Fisher Scoring iterations: 6
```

Output 2. Stepwise Reduced Model

```
glm(formula = sale ~ recmon + ordcls2 + ordcls3 + ordcls4 + ordcls5 +
    salcls1 + salcls3 + salcls5 + salcls6 + ord185 + ord285 +
    ord385 + ord485 + totsale + freq.buyer, family = binomial,
    data = traintrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3116  -0.2865  -0.2390  -0.2022   2.9432

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.4159734  0.0732880 -46.610  < 2e-16 ***
recmon         -0.0145242  0.0025790  -5.632 1.78e-08 ***
ordcls2         0.0877071  0.0444924   1.971 0.048691 *
ordcls3         0.2246048  0.0820187   2.738 0.006173 **
ordcls4         0.3553126  0.1675171   2.121 0.033917 *
ordcls5        -0.1837017  0.0913244  -2.012 0.044270 *
salcls1        -0.0023411  0.0015267  -1.533 0.125169
salcls3         0.0022808  0.0004755   4.797 1.61e-06 ***
salcls5         0.0025735  0.0012416   2.073 0.038196 *
salcls6         0.0017516  0.0004690   3.735 0.000188 ***
ord1851         1.1760357  0.1032020  11.395  < 2e-16 ***
ord2851         0.6125092  0.1053317   5.815 6.06e-09 ***
ord3851         0.8261934  0.0710256  11.632  < 2e-16 ***
ord4851         0.7771072  0.0881062   8.820  < 2e-16 ***
totsale        -0.0001461  0.0001016  -1.438 0.150428
freq.buyerTRUE  0.1475279  0.0632113   2.334 0.019602 *
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16273  on 39999  degrees of freedom
Residual deviance: 13956  on 39984  degrees of freedom
AIC: 13988
```

Output 3. Further Reduced Model

```
Call:
glm(formula = sale ~ ord185 + ord285 + ord385 + ord485 + freq.buyer,
    family = binomial, data = traintrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5813  -0.2651  -0.2154  -0.2154   2.7479

Coefficients:
                Estimate Std. Error  z value Pr(>|z|)
(Intercept)    -3.75229    0.03631 -103.331  < 2e-16 ***
ord1851         1.74382    0.04959   35.162  < 2e-16 ***
ord2851         1.03984    0.06072   17.124  < 2e-16 ***
ord3851         0.79556    0.06874   11.573  < 2e-16 ***
ord4851         0.66439    0.07994    8.311  < 2e-16 ***
freq.buyerTRUE  0.42136    0.05238    8.044 8.71e-16 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16273  on 39999  degrees of freedom
Residual deviance: 14074  on 39994  degrees of freedom
AIC: 14086

Number of Fisher Scoring iterations: 6
```

Output 4. Reduced Model with Interactions

```
Call:
glm(formula = sale ~ ord185 + ord285 + ord385 + ord485 + freq.buyer +
    ordcls3:recmon + ordcls3:tof + ordcls3:salcls2 + ordcls3:salcls3 +
    ord285:ordcls3 + ord485:ordcls3 + ordcls3:totsale, family = binomial,
    data = traintrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8356  -0.2613  -0.2093  -0.2093   2.7685

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     -3.810e+00  3.995e-02 -95.383  < 2e-16 ***
ord1851          1.575e+00  6.504e-02  24.211  < 2e-16 ***
ord2851          1.169e+00  1.693e-01   6.902 5.12e-12 ***
ord3851          8.168e-01  6.964e-02  11.729  < 2e-16 ***
ord4851          8.440e-01  1.031e-01   8.185 2.72e-16 ***
freq.buyerTRUE   4.500e-01  5.503e-02   8.177 2.92e-16 ***
ordcls3:recmon   1.307e-02  5.124e-03   2.550 0.010759 *
ordcls3:tof      1.324e-03  3.596e-04   3.681 0.000232 ***
ordcls3:salcls2  1.330e-03  4.866e-04   2.734 0.006256 **
ordcls3:salcls3  1.322e-03  2.614e-04   5.056 4.28e-07 ***
ord2851:ordcls3 -3.371e-01  9.441e-02  -3.570 0.000357 ***
ord4851:ordcls3 -2.258e-01  8.743e-02  -2.582 0.009815 **
ordcls3:totsale -2.779e-04  8.978e-05  -3.095 0.001965 **
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16273  on 39999  degrees of freedom
Residual deviance: 14010  on 39987  degrees of freedom
AIC: 14036

Number of Fisher Scoring iterations: 6
```

Output 5. Multiple Regression Model 1

```
Call:
lm(formula = targamnt_log ~ ordcls1 + ordcls3 + ordcls7 + AOA_log +
    AOA2_log + AOA3_log + AOA5_log + AOA6_log + AOA7_log + ord185 +
    ord285 + ord485 + tof + totord + totsale + PR, data = train4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.66723 -0.43259 -0.03444  0.40255  1.76436

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.859e+00  1.116e-01  25.614  < 2e-16 ***
ordcls1     -1.485e-01  4.676e-02  -3.177 0.001506 **
ordcls3      1.176e-01  2.955e-02   3.979 7.10e-05 ***
ordcls7     -1.192e-01  4.374e-02  -2.726 0.006457 **
AOA_log      3.043e-01  2.706e-02  11.242  < 2e-16 ***
AOA2_log    -3.650e-02  7.699e-03  -4.741 2.24e-06 ***
AOA3_log     4.432e-02  1.097e-02   4.040 5.50e-05 ***
AOA5_log     1.797e-02  1.069e-02   1.680 0.093058 .
AOA6_log    -2.325e-02  7.598e-03  -3.060 0.002233 **
AOA7_log     4.476e-02  1.869e-02   2.395 0.016703 *
ord185      -2.060e-01  4.725e-02  -4.359 1.35e-05 ***
ord285      -1.704e-01  4.481e-02  -3.804 0.000146 ***
ord485      -7.109e-02  3.419e-02  -2.080 0.037660 *
tof          7.968e-04  2.763e-04   2.884 0.003959 **
totord      -1.688e-02  3.213e-03  -5.253 1.61e-07 ***
totsale      6.283e-04  6.365e-05   9.871  < 2e-16 ***
PR           5.263e-01  3.064e-01   1.718 0.085972 .
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

Residual standard error: 0.5992 on 2792 degrees of freedom
Multiple R-squared:  0.2088,    Adjusted R-squared:  0.2043
F-statistic: 46.06 on 16 and 2792 DF,  p-value: < 2.2e-16
```

Output 6. Multiple Regression Model 2

```
Call:
lm(formula = targamnt_log ~ recmon + salcls1_log + salcls2_log +
    salcls3_log + salcls5_log + breadth + ord185 + ord285 + ord485 +
    totord + totsale + PR, data = train4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.73294 -0.43690 -0.04068  0.42228  1.79384

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.980e+00  4.421e-02  90.026  < 2e-16 ***
recmon       2.239e-03  1.349e-03   1.660  0.09702 .
salcls1_log -3.151e-02  1.457e-02  -2.163  0.03063 *
salcls2_log -1.460e-02  8.293e-03  -1.761  0.07841 .
salcls3_log  1.002e-01  1.200e-02   8.353  < 2e-16 ***
salcls5_log  2.936e-02  1.113e-02   2.637  0.00842 **
breadth     -2.620e-02  1.170e-02  -2.239  0.02523 *
ord185      -1.766e-01  4.391e-02  -4.021 5.95e-05 ***
ord285      -1.533e-01  3.918e-02  -3.914 9.31e-05 ***
ord485      -7.761e-02  3.472e-02  -2.235  0.02549 *
totord      -2.807e-02  2.256e-03 -12.446  < 2e-16 ***
totsale      9.882e-04  5.503e-05  17.959  < 2e-16 ***
PR           4.866e-01  2.531e-01   1.923  0.05460 .
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

Residual standard error: 0.6122 on 2797 degrees of freedom
```

```
Multiple R-squared:  0.1704,    Adjusted R-squared:  0.1669
F-statistic: 47.88 on 12 and 2797 DF,  p-value: < 2.2e-16
```

Output 7. Multiple Regression Model 3

```
Call:
lm(formula = targamnt_log ~ salcls1_log + salcls2_log + salcls6_log +
    prefer1 + ord185 + ord285 + ord385 + ord485 + totord + totsale +
    PR, data = train4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.72289 -0.44078 -0.05098  0.42196  1.83466

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.111788   0.024006 171.280  < 2e-16 ***
salcls1_log -0.082341   0.022169  -3.714 0.000208 ***
salcls2_log -0.034151   0.007462  -4.576 4.94e-06 ***
salcls6_log -0.020592   0.007353  -2.800 0.005141 **
prefer1      0.259991   0.115329   2.254 0.024251 *
ord185       0.059398   0.025619   2.319 0.020492 *
ord285       0.044234   0.027040   1.636 0.101977
ord385      -0.049193   0.032083  -1.533 0.125317
ord485      -0.085020   0.035368  -2.404 0.016287 *
totord      -0.032079   0.002224 -14.424  < 2e-16 ***
totsale      0.001135   0.000055  20.644  < 2e-16 ***
PR           0.405224   0.241038   1.681 0.092843 .
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

Residual standard error: 0.6186 on 2798 degrees of freedom
Multiple R-squared:  0.1556,    Adjusted R-squared:  0.1523
F-statistic: 46.88 on 11 and 2798 DF,  p-value: < 2.2e-16
```

Output 8. Multiple Regression Model 4

```
Call:
lm(formula = targamnt_log ~ ordcls1 + ordcls3 + ordcls7 + AOA_log +
    AOA2_log + AOA3_log + AOA5_log + AOA6_log + AOA7_log + ord185 +
    ord285 + ord485 + tof + totord + totsale + PR + AOA6_log:tof,
    data = train4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.68267 -0.42758 -0.03489  0.40169  1.77351

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.854e+00  1.117e-01  25.558  < 2e-16 ***
ordcls1      -1.493e-01  4.675e-02  -3.192 0.001427 **
ordcls3       1.154e-01  2.958e-02   3.900 9.86e-05 ***
ordcls7      -1.220e-01  4.377e-02  -2.788 0.005343 **
AOA_log       3.059e-01  2.708e-02  11.296  < 2e-16 ***
AOA2_log     -3.759e-02  7.732e-03  -4.861 1.23e-06 ***
AOA3_log      4.285e-02  1.101e-02   3.890 0.000102 ***
AOA5_log      1.822e-02  1.069e-02   1.704 0.088521 .
AOA6_log     -3.466e-02  1.079e-02  -3.212 0.001333 **
AOA7_log      4.462e-02  1.869e-02   2.388 0.017019 *
ord185       -2.050e-01  4.725e-02  -4.339 1.48e-05 ***
ord285       -1.651e-01  4.494e-02  -3.673 0.000244 ***
ord485       -6.891e-02  3.421e-02  -2.014 0.044076 *
tof           7.708e-04  2.768e-04   2.785 0.005391 **
totord       -1.800e-02  3.300e-03  -5.456 5.30e-08 ***
totsale       6.281e-04  6.364e-05   9.870  < 2e-16 ***
PR            6.671e-01  3.206e-01   2.081 0.037549 *
```

```
AOA6_log:tof  1.265e-04  8.501e-05   1.488 0.136763
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

Residual standard error: 0.599 on 2791 degrees of freedom
Multiple R-squared:  0.2095,    Adjusted R-squared:  0.2046
F-statistic:  43.5 on 17 and 2791 DF,  p-value: < 2.2e-16
```

Output 9. Multiple Regression Model 5

```
Call:
lm(formula = targamnt_log ~ ordcls1 + ordcls3 + ordcls7 + AOA_log +
    AOA2_log + AOA3_log + AOA5_log + AOA6_log + AOA7_log + ord185 +
    ord285 + ord485 + tof + totord + totsale + AOA3_log:PR, data = train4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.72552 -0.43151 -0.03401  0.40208  1.76477

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.896e+00  1.075e-01  26.927  < 2e-16 ***
ordcls1     -1.536e-01  4.667e-02  -3.291 0.001011 **
ordcls3      1.001e-01  3.002e-02   3.333 0.000869 ***
ordcls7     -1.140e-01  4.362e-02  -2.615 0.008980 **
AOA_log      3.063e-01  2.701e-02  11.337  < 2e-16 ***
AOA2_log    -3.488e-02  7.380e-03  -4.727 2.39e-06 ***
AOA3_log     2.220e-02  1.260e-02   1.762 0.078232 .
AOA5_log     1.816e-02  1.065e-02   1.705 0.088368 .
AOA6_log    -2.378e-02  7.312e-03  -3.252 0.001159 **
AOA7_log     4.402e-02  1.864e-02   2.362 0.018256 *
ord185      -2.025e-01  4.716e-02  -4.294 1.82e-05 ***
ord285      -1.514e-01  4.519e-02  -3.351 0.000817 ***
ord485      -7.159e-02  3.413e-02  -2.097 0.036053 *
tof          8.868e-04  2.464e-04   3.599 0.000325 ***
totord      -1.742e-02  3.039e-03  -5.733 1.09e-08 ***
totsale      6.071e-04  6.396e-05   9.492  < 2e-16 ***
AOA3_log:PR  2.804e-01  8.263e-02   3.393 0.000700 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1

Residual standard error: 0.5982 on 2792 degrees of freedom
Multiple R-squared:  0.2112,    Adjusted R-squared:  0.2067
F-statistic: 46.73 on 16 and 2792 DF,  p-value: < 2.2e-16
```

References

[1] Tamhane A.C. and Malthouse E.C., *Predictive Analytics (I): Parametric Regression and Classification Models*. Book draft, Wiley Interscience. 2015.

[2] Chatterjee S., Hadi A.S., *Regression Analysis by Example*, 5th ed. Hoboken: Wiley. 2012.

[3] Powers D.M.W. "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation." *Journal of Machine Learning Technologies* vol. 2, iss. 1 (2011):37-63. Available `http://www.flinders.edu.au/science_engineering/fms/School-CSEM/publications/tech_reps-research_artfcts/TRRA_2007.pdf`

[4] Porzak, J. "Using R for Customer Segmentation." Presentation at useR! conference. Dortmund, Germany. August 2008. Available `https://ds4ci.files.wordpress.com/2013/09/user08_jimp_custseg_revnov08.pdf`

[5] Domencich, T., McFadden, D. L. "Statistical Estimation of Choice Probability Functions." In *Urban Travel Demand: A Behavioral Analysis*, 101-125. New York: American Elsevier Publishing Co. 1975. Available `http://eml.berkeley.edu/~mcfadden/travel/ch5.pdf`