**Case Assignment: CD Now**

**Part 1: Traditional RFM: Decile Analysis**

|  |
| --- |
| **Decile vs Percent Retained** |
|  |
|  |
|  |

Recency and monetary seem to make a difference in retention. In deciles 1 through 4, there are varied retention rates. With those who have shopped recently having higher rates of retention than those who have not. Deciles 5 through 10 have similar retention rates. Each monetary decile has very different retention rates. Both monetary and recency suggest that there are differences in retention rates between groups depending on how recently a customer shopped and how much they spent. When attempting these analyses for frequency, we see that there is little difference between retention rates in each decile. From this analysis, it seems that frequency does not have a significant effect on retention. It should be noted that frequency is very related to recency.; the more often someone visits a store, the fewer time spent between visits.

**Part 2: Visualizations with Cumulative Lift Charts**

|  |  |
| --- | --- |
| **Cumulative Lift Charts** | |
|  |  |

Based on these charts, recency appears to be more important for predicting retention. The cumulative lift chart for recency shows that there is 2.5 times more predictive value than the overall retention rate. Using the monetary cumulative lift chart, we see that targeting the first decile only produces a 1.16 lift over the overall response rate, it is not as steep as the recency lift chart, and retention rate increases in deciles, suggesting that customers who spent less are better retained as we move into other deciles, which would not be the case if the monetary model was a good model. A good model would have 100% retention rate in the first decile, and declining retention rates thereafter.

**Part 3: Predicting Retention with Regression**

**Linear Regression**

From the linear regression, there is a significant effect of how recently someone has bought merchandise, how frequently they buy merchandise, and the monetary value of their order on retention. That is to say, the more recently, the more often, and the more one spends on merchandise at CD Now the more likely they are to be retained as a customer. The more someone engages with CD Now in meaningful ways, the more likely they are to be retained.

|  |  |
| --- | --- |
|  |  |
|  |  |

The strengths of linear regression include its ability to control for other variables, demonstrate statistical significance, and make predictions for future customers. However, linear regression is weak in this instance because it predicts values from – infinity to + infinity. In this case, we want to estimate probability, i.e. we need predictions ranging from 0 to 1. Any predictions the linear regression makes outside of that range are uninterpretable.

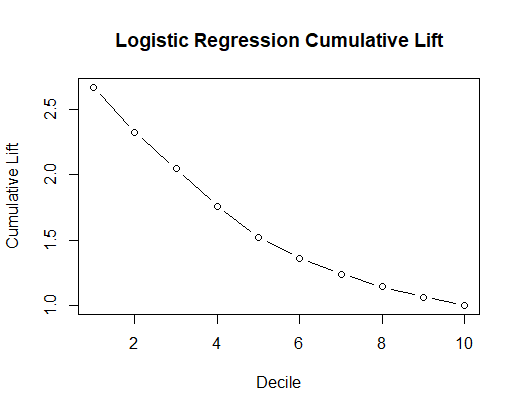
**Logistic Regression**

Again, we see the same results for the logistic regression, there is a significant effect of how recently someone has bought merchandise, how frequently they buy merchandise, and the monetary value of their order on retention. In other words, the more recently, the more often, and the more one spends on merchandise at CD Now the higher probability they have of being retained.

|  |  |
| --- | --- |
|  |  |
|  |  |

The logistic model has many of the same strengths that the linear regression possesses, such as controlling for variables, demonstrating statistical significance, and making predictions about future customers. However, it has a weakness in its interpretation of coefficients (especially compared to the coefficients of the linear regression) because of the exponentiation the model underwent. The logistic model is preferable because it is a probability classifier while the linear regression predicts values between -infinity and + infinity. This gives us predictions between 0 and 1, which can be interpreted as probabilities of being retained. The models lead to the same broad conclusions, that recency, frequency, and monetary have significant main effects on retention.

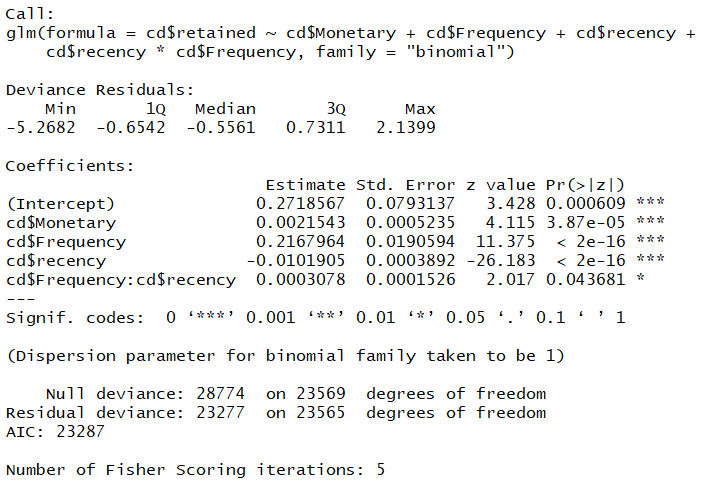
**Part 4: Visualizing Logistic Regression with a Lift Chart**



I conclude, that the logistic model is a better predictor of retention rates than any of the pieces alone from the traditional RFM analysis. To begin, in decile 1, there is 2.6 boost in predictive value over the overall response rate. We capture a higher percentage of respondents in the first decile; fewer people may be targeted and achieve better results. As we move along the x-axis, each decile captures fewer people, meaning that we are accurately predicting retained customers with the model. The cumulative lift chart is steeper than the others also. The logistic model is superior to the RFM analysis because of its ability to handle continuous data, control, and make predictions for future customers.

**Part 5: Interaction Modeling and Cross Validation**

**Interaction Modeling**



There is a significant interaction between frequency and recency. There is a significant interaction of frequency on retention that is dependent on recency. Frequency customers who have shopped more recently have higher retention rates. Those who have shopped more often are more likely to be retained if they made a purchase recently.

**Cross Validation**

|  |  |
| --- | --- |
| **Model** | **5-Fold CV Results** |
| Model with Interaction | 1.66 |
| Model without Interaction | 0.40 |

The cross validation shows that both models have a low rate of predicted error via the induced models. Meaning, both models make accurate predictions. Based on the cross-validation, the model without the interaction is better. In this case, to be better means correctly guessing the number of true positives and true negatives, to correctly guess the customers who are retained and those who have left.

**Part 6: Conclusions and Managerial Implications**

From the analysis, it appears that retention is dependent on recency and frequency. Because of the relatedness of the two variables, I think CDNow management should focus on bringing customers in more often, also meaning that customers have less time between purchases (they are more recent). It is better to have small marginal gains on customer, by encouraging them to visit more often and purchase. This could have helped CDNow’s longevity and increase the product velocity in store shelves.