Relay Food Management Case Study: Homework 2
MAR 653 Marketing Analytics

Group 3
Sana Khan
Jessica Woolson
Robert Femia
Caitlyn Hollowell

May 16, 2024

EXECUTIVE SUMMARY

How do we help Retail Relay predict churn to ultimately retain customers and grow

Challenge

Retail Relay an online grocery service has non-contractual agreements with it's customers, which makes it difficult to know if a customer has churned. We will help Retail understand what predictors show if a customer is retained or churned

Analysis

We ran multiple logistic models to understand what variables could predict retention

- **Multiple variables were statically significant in predicting customer retention**
- **The model with the highest hit rate included a dummy variable and had a hit rate of 94%**

Next Steps

- **Incentivize customers to sign up for automatic refills and doorstep delivery**
- **Send emails that are segmented to customers and increase click rate**
- **Have customers sign up for paperless delivery**

Syracuse University

THE CHALLENGE

Retail Relay, was created with the goal of having customers save time and money by providing a a platform where they could order all their groceries from multiple vendors and have one pick up spot.

Despite being an innovate retailer, they faced significant challenges with customer retention. A pilot study was done to understand customer behavior and the team noticed that 45% of customers did not make a second purchase. Because Retail Relay has a non-contractual relationship with their customers, they had to use p(Alive), a probability metric to estimate the likelihood that a customer is still active or "alive."

The management team wanted to assess customer activity and analyze data on customer relationships to identify what characteristics retained a customer. The goal is to use the information from the customer data to make the changes necessary to retain customers.

USING LOGISTIC REGRESSION MODELS TO PREDICT CUSTOMER RETENTION

Multiple logistic regression models were created to test the best fit for predicting the probability of customer retention using the relay test data set

Independent (x) Predictor Variables Used	Model 1	Model 2	Model 3	Model 4
essent	x		x	x
eclickrate	x			x
avgorder	x	x		x
ordfreq	x	x		x
paperless	x	x		x
refill	x	x		x
doorstep	x	x		x
weekend				x
Accurate observations in model	5847	4933	5752	5856
Total observations in model	6219	6219	6219	6219
Hit rate	94.02%	79.32%	92.49%	94.16%

Model #4, using all predictor variables including the 'weekend' dummy variable (fav day was Fri/Sat/Sun) produced the **highest hit and accuracy rate** in the test data set

Esent Predictor Variable

The emails sent predictor variable has a coefficient of 0.2115 in the previous model and its p-value is significant as it is less than 0.1. In comparison to other predictor variables in the model, esent doesn't have as large of an impact on the predicted probability of customer retention.

Issues w/ Using Esent

Issues with using the number of emails sent to a customer for their predicted retention with Relay lies in a few different possible factors:

- the number of emails sent to a customer only takes into account volume of sends vs. the amount of engagement with a particular customer
- The number of emails sent by Relay has multi-collinearity (interdependence) with other predictor variables that could be used in the regression model.
    - For example esent could also be related to eclickrate, ordfreq, paperless, etc. which will influence the results and confidence in the model

Addressing Potential Issues w/ Esent

If esent is used as a predictor it could be possible to run a log transformation to normalize the data and reduce the impact of outliers. It could also be possible to remove essent as a variable altogether. Alternative variables such as eclickrate could be used. Eclick rate is Number of emails clicked divided by number of emails sent which shows the rate that customers are opening emails.

Syracuse University

USING LOGISTIC REGRESSION MODELS TO PREDICT CUSTOMER RETENTION

MODEL OUTPUTS

1

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.5348036  0.0777869 -32.587  < 2e-16 ***
esent        0.2110782  0.0032626  64.696  < 2e-16 ***
eclickrate   0.0192004  0.0017564  10.932  < 2e-16 ***
avgorder    -0.0035576  0.0007572  -4.698 2.62e-06 ***
ordfreq     -0.6074376  0.2313531  -2.626  0.00865 **
paperless    0.5252447  0.0641304   8.190 2.61e-16 ***
refill       0.7693665  0.1040461   7.394 1.42e-13 ***
doorstep     0.8659042  0.1602369   5.404 6.52e-08 ***
```

2

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.668536   0.039065  17.113  < 2e-16 ***
avgorder    0.002025   0.000453   4.469 7.85e-06 ***
ordfreq     0.073195   0.156046   0.469    0.639
paperless   0.809074   0.033121  24.428  < 2e-16 ***
refill      0.867731   0.079470  10.919  < 2e-16 ***
doorstep    0.851283   0.131893   6.454 1.09e-10 ***
```

3

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.117615   0.042094  -50.31   <2e-16 ***
esent        0.211453   0.003121   67.76   <2e-16 ***
```

4

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.6036352  0.0797664 -32.641  < 2e-16 ***
esent        0.2115356  0.0032734  64.622  < 2e-16 ***
eclickrate   0.0192718  0.0017602  10.949  < 2e-16 ***
avgorder    -0.0034973  0.0007535  -4.641 3.46e-06 ***
ordfreq     -0.5967557  0.2313996  -2.579  0.00991 **
paperless    0.5192569  0.0642167   8.086 6.17e-16 ***
refill       0.7607544  0.1040587   7.311 2.66e-13 ***
doorstep     0.8471091  0.1606152   5.274 1.33e-07 ***
weekend      0.2640833  0.0654575   4.034 5.47e-05 ***
```

Directional Interpretation of Coefficients for Model 4

Avgorder

- the average order size of the customer shows a negative impact on the predicted probability of customer retention
- the sign on this coefficient isn't as intuitive, as we might naturally assume that customers who purchase more on average at Relay are more loyal to the store and therefore and predicted at a higher rate to stay with the company as a customer
- a possible explanation for this trend could be that customers who make one-off larger customers with the store aren't as likely to remain a customer while the majority of their customer base makes smaller purchases (perhaps more frequently) and is more loyal to Relay. Another possibility is that customers who make larger orders might have higher expectations and if they are disappointed they may churn

Ordfreq

- the number of orders divided by customer tenure (ordfreq) shows a negative impact on the predicted probability of customer retention
- it is not intuitive that having more orders would be associated with a reduced probability for retention since you'd this more orders = more loyal
- A possible explanation for this is frequent ordering may be a result of not getting the correct product, which can lead to dissatisfaction and churn

Weekend

- whether or not the favday is a Friday, Saturday or Sunday shows a positive impact on the predicted probability of customer retention
- the sign on this coefficient is more intuitive. this positive impact might suggest that customers who are choosing their favorite delivery day as Friday, Saturday or Sunday are likely to have increased probability for customer retention.
- weekend activity might reflect leisure time usage, indicating customers find the service convenient and worthwhile during their free time.

Syracuse University

NEXT STEPS / IMPROVEMENT FOR CUSTOMER RETENTION

Next Steps

- Create a customer loyalty program
- Advertise on social media (Facebook, Instagram)
- Paid advertisement campaign on various platforms (YouTube, Amazon) that focus local, high click-rate areas
- Source new vendors to include on the website
- Provide incentives to increase adoption of doorstep delivery subscription and automatic refills – as these are variables with higher impact for retention probability

Improvements

- Create a customer loyalty program tailored to frequent customers and that will attract new ones
- Solicit customer feedback to identify pain-points and make improvements
- Send emails to customers who have been dormant for certain period (deals/discounts)
- Increase marketing activities on weekends and offer special promotions to attract and retain customers (based on high-activity days model)
- Prompts/inquiries for paperless notification option – as this is a variable with high impact on probability for retention
- Add more vendors so customers have complete place to shop


Syracuse University
Whitman School
of Management