Greyson Corporation Marketing Department

MSBA Spring 2021 Case Competition

Renewal Rate Analysis

Nirosha Telu, Mia Simms, Marshall Ferguson,

And Christopher Walker, Asaad Khattab

**Contents:**

1. Case Background
2. Data Preparation and Understanding
3. Modeling
4. Our Actionable Insights to Greyson Corp.,
5. Reference

**Case Background**

Greyson Corporation is a media and marketing company involved in magazine and book publishing and broadcasting. Greyson also provides of a range of services including market research, communications planning, and website advertising. that can enhance their client’s brands. Greyson uses data to develop campaigns for new customer acquisition, customer reactivation, and the identification of cross-selling opportunities for products.

Greyson Corp. will generate separate versions of a monthly issue of a magazine that will differ only by the advertisements they contain. They will mail a subscribing customer the version with the print ads identified by their data as being of most interest to that customer. Greyson Corporation reached out for help with the analytics behind customer renewal rate, to determine how to boost efforts in increasing market reach for not only themselves but their clients.

The goal to come up with a marketing strategy to boost customer response to the company’s renewal offers to increase customer retention. Currently, the company will mail renewal offers to their magazine subscribers. The industry response rate is around 2%. However, Greyson Corporation has historically outperformed this average and currently has a response rate of 2.04%. The company must update their model to work with recent changes. The director of marketing has a goal of ensuring that the Greyson Corporation maintains its place as one of the top players in targeted marketing.

This is a supervised machine learning problem dealing with classifying whether or not a customer will respond to Greyson Corporation’s subscription offer. We will attempt to determine the variables and factors that will predict whether or not a customer will respond to a subscription offer, and how strongly these variables and factors affect the rates of subscription offer responses.

**Data Preparation and Understanding**

The Greyson dataset consists of 36,000 observations and 36 predictor variables. There are 38 total variables in the set including the unique identifier for each customer, the Customer ID. The Renewal variable is the response variable of the dataset. It takes two values, 1 for those who renew and 0 for those who do not.

Several variables focused on demographic data for the subscribers. The age variable only took values 18+ since only adults can subscribe. The most frequent age for subscribers was consistent with the mean and median values for this variable, around 52 years old. Gender, when known, it was also noted, using historical labels of male and female. The majority of the subscribers, over 29,000, are female. Marital Status was listed as single, married, previously married, or unknown. The majority of the subscribers are married (26,095 observations). The average income was nearly $91,000, while the median. The highest income observed is $500,000, the lowest was $5,000.

The subscriber’s Occupation is also noted in general categories. There are classifications for retirees, homemakers, and unknown occupations. The additional classifications were quite broad for white-collar, blue-collar or middle-class jobs. About 35% of the subscribers hold white-collar job titles like executive.

Most subscribers live in single-family homes, over multi-family homes and apartments. Home value is categorized into $50,000 intervals from $0-$400,000. However, these ranges are indicated by one-to-two-digit integers. The higher the value in the data set, the higher the value of the home. For example, a home with a rating of 4 is valued at $150,000 to $199,999. A home is rated at 5 is valued between $200,000 to $249,999. The rating of 10 is a bit different, it’s used to categorize a home with unknown value.

The dataset considers the likelihood of a customer being a homeowner. High values reflect the higher the chance of the subscriber owning the home they live in. The average value for this variable is 89, which would be a high likelihood of ownership. The most frequent value is 100, learning most individuals own their homes who have subscribed to Greyson. Roughly 86% of the subscribers are confirmed, homeowners. They also establish the length of time that the subscriber has lived at the current address. Using 1 for anything less than two years and going up to 14 years or more. The dataset includes information about household size categorizing them from 1 to 6 or more members in the household. The data notes whether children are present at the residence and then groups on the likelihood of the children being between ages 0-5, 6-12 or 13-18. The data dictionary categorizes children as anyone 21 or younger though. This magazine appeals to people across various socio-economic backgrounds, but mainly upper-middle-class, married women, working white collar jobs who are homeowners living in single family home settings.

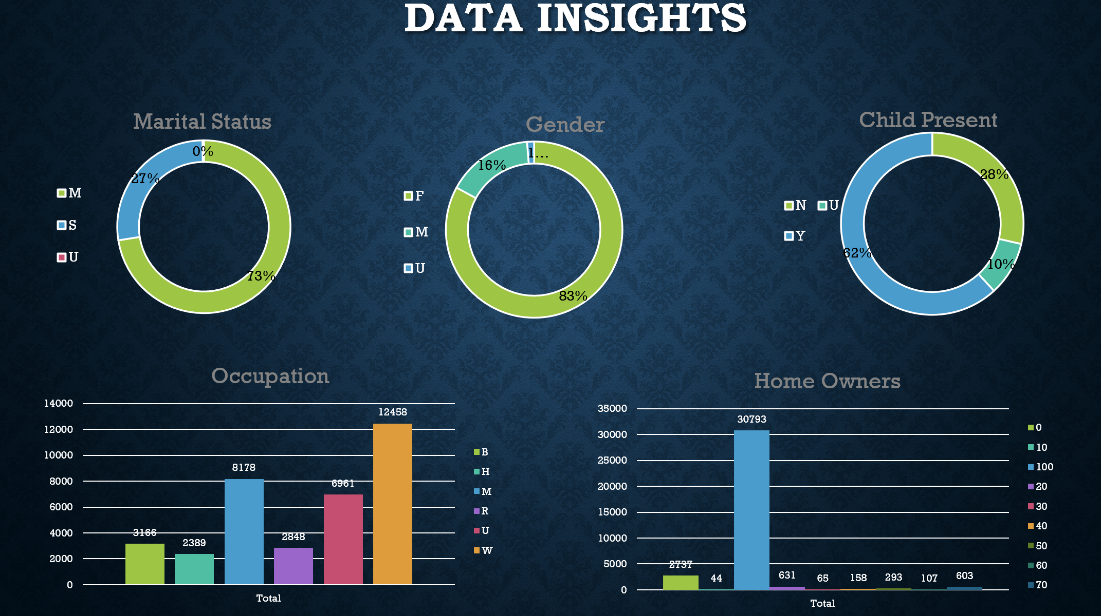
The remaining variables are concentrated on status and statistics associated with the subscribers’ subscription with Greyson Corp. The Magazine status variables indicate the customer’s subscription status. Some of the classifications include whether the subscription is active, cancelled, expired, gifted, or suspended. There is also a Magazine Status category designated to subscriptions where the current status is unknown and that indicates, how long a subscription has been expired or cancelled. There is a variable that totals the number of paid direct mail orders Greyson Corp has from customers across all their magazine subscriptions, one for the number of years since the last order was made across all their business lines. Greyson Corp tracks the total amount of money paid for all their magazine subscriptions overtime per customer.

There are also aggregated variables that record the average value per issue and the total number of orders paid per customer across all magazines offered. Greyson Corp records the number of magazine titles in the following categories: unpaid status, paid cash, paid to reinstate, paid credit, inactive status, expired status, cancelled for non-payment, and in paid complaint status. These are useful in identifying how many magazines are important to each individual subscriber and how engaged they are with their subscriptions. The values taken by these variables could be as low as 0 and as high as the total number of magazines offered by Greyson Corporation.

Greyson tracks the number of months since the most recent payment and the payment method that customers used on their most recent order. Subscribers can pay with a gift card, gift billed to another payer, cash credit, or via an online portal. There is also financing available to cover the cost of the subscription. This variable also indicates whether the customer requested cancellation, advanced renewal order cancellations, or if payment was received after the order cancelled. There are also categories in this variable for defaulted accounts, no billing accounts and accounts covered by paid credit. There is a separate category that indicates whether the account is the result of a gift from a donor account and another the tracks the number of gift orders per customer. There are also separate variables to keep track of the months since the customers’ first order, most recent order and the number of months since the subscriptions has expired.

One of the biggest inconsistencies found in the dataset is based on the way children are defined in the Child Present Variable as anyone younger than 22 years of age but the age groups for the categorical variables only cover children ages 0-18. A category to account for ‘children’ 19-21 should be added for the analysis to make sure that no factor is being overlooked.

There are several variables in the data that are missing values, including the likelihood of homeownership (569). This may be hard to determine without the customer’s direct input. The other variables that include missing values are Dollar Per Issue, Total Orders Paid, Months Since Last Payment and Months Since Expired. Dollar Per issue helps identify the value per magazine which may prove important to consumers when deciding whether the subscriptions worth keeping. All remaining missing data regarding order cost could be useful in making more accurate predictions. TAC will run all modelling with and without missing values to see which more closely fits the data.



Overall, the dataset consists of 8 categorical variable and 30 numeric variables. The target variable, “Renewal”, has 35,263 subscribers that did not renew their magazine in response to a mailing, and 737 or 2.05% of the subscribers that did renew their magazine in response to a mailing. There are 29,849 female subscribers, 5,689 male subscribers, and the gender is unspecified for 462 subscribers.

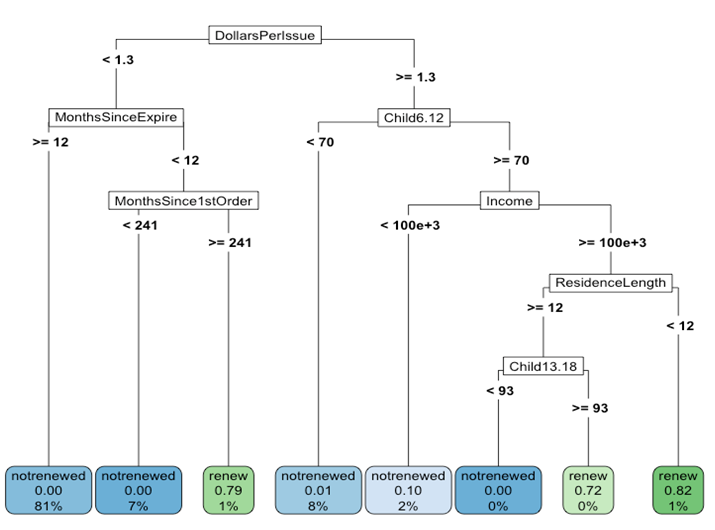
The highly correlated factors with whether a subscriber responds to mailing or not are the subscriber’s income, the likelihood of child 6-12 is present in the home, the likelihood of child 13-18 is present in the home, the total amount paid for all magazine subscriptions over time, average value per issue, and frequency (in months) of 1st order for the magazine.

**Modeling**

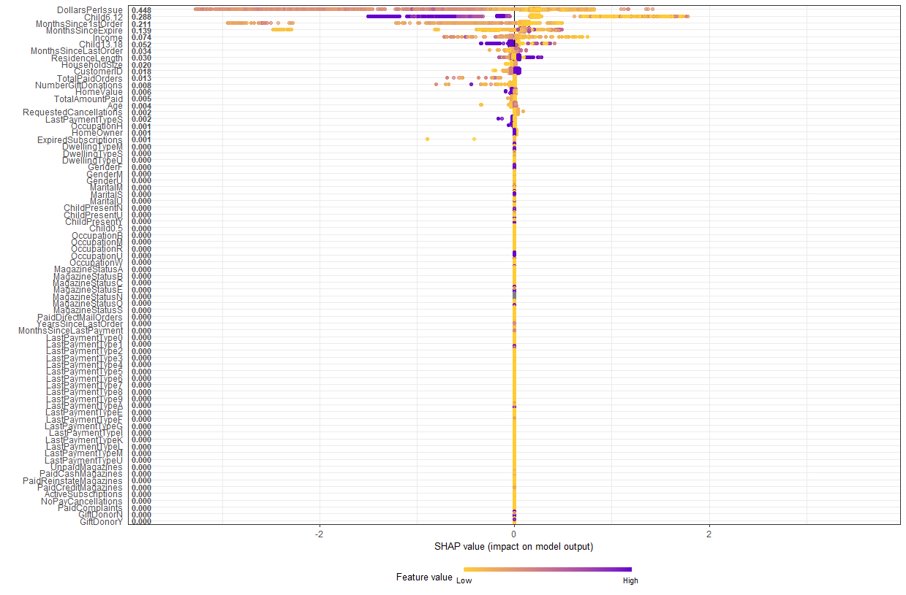
Initially four classification machine learning algorithms were tested against the dataset: Lasso Logistic Regression, Random Forest, Decision Tree, and Gradient Boosting Trees. Before modeling, the data between training set with 80% of the observations and a test set with 20 % of the observations. The median values of variables with missing values were used to impute those missing values in both the training and testing sets after the partition. Five-fold cross-validation was used to determine which of the five subgroups of the training set performed better on new data.

Each model was ran using the same seed number to make the results were comparable as possible. During the evaluation process to decide which model would best help Greyson, who would using the model within Greyson Corporation was also consider. Due to that and it’s high AUC value, the decision tree model was selected as the premier model to represent customer renewal rate response. Trees can be visualized and are easy to interpret and be used for prediction. Since a sales team who may not have significant understanding of more complicated models who be who uses this tool within Greyson, interpretability, by non-technical team members, was of the most importance. The visual aspect of this models would help convey the findings and any potential insights to the non-technical audience. The way a decision tree works is by separating the different variables in the dataset into subgroups. The different subgroups help predict different outcomes of our dependent variable, in this case, whether a customer will renew.

This model also identifies pain points for customers who would choose to not renew. It focuses on the variables that were considered most impactful on the response variable’s outcome. The decision trees put the most impactful variables at junctions and create paths that split based on values that is the most impactful on a customer’s decision to renew or not. These pathways lead to the actually expected response for future customers.



The Shapley plot will help Greyson Corporation better understand the impact of the values of each variable and how it affects whether or not a customer renews their subscription. It will be useful in building recommendations. In the plot below, that higher cost (in dollars) per issue will likely make a customer churn. When there are kids 6-12 in the home, customers are more likely to churn. Subscribers without kids 6-12 years old are more likely to renew. The higher the number of months since the first order, the lower the likelihood of a customer renewing. These are the three most impact variables on a customer’s decision to renew.



**Our Actionable Insights to Greyson Corp.,**

As Business Analysts, we have a good understanding of the business problem that Greyson Corp is facing and as a part of exploring their historic data, running various analysis on the data helped us to recommend few crucial and actionable steps to help Greyson to reach their goal of leading the industry response rate and boost the renewal subscription rate.

The marketing strategy that we could build on Greyson data is the STP approach and this explains the necessary actionable insights that they can take.

**Recommendations using STP Approach:**

As Greyson wanted us to identify the targeted groups of potential subscribers and non-subscribers, we have identified those groups by this approach and STP stands for **S**egmentation of Market-**T**argeting valued groups-**P**ositioning the offering to attract customers to new subscriptions and retain the old customers.

With segmenting the market by observing the demographics of the data and by identifying the targeted groups in the Market, we suggest Greyson follow the following insights.

As the dollar per issue is one of the most important variables that is impacting the renewable rate of the magazine subscriptions, we would like to suggest that Greyson should be focusing more on the cost per issue of the magazine and add a value proposition of the magazine subscriptions by initiating offers or coupons that targets the most groups which are likely to subscribe. Offer a coupon toward the end of the subscription that includes one free magazine for renewing, which lowers the cost per issue. This would make customers feel more inclined to renew.

We also identified that there are families who got subscribed to the 1st order of magazine and it is very important to Greyson to note that as the renewing months since 1st order of magazine increases, the customers are less likely to renew. In order to fill this gap, we suggest Greyson does one or more of the following: require automatic subscriptions that must be canceled rather than requiring customers to complete something to renewal, or send an enticing renewal offer within the first three months of subscribers having the magazine subscription. The second is less preferable as we don’t currently know when Greyson begins to request renewals. However, the data indicates the earlier subscribers are approached to renew, the higher the likelihood they actually renew.

We recommend Greyson for partner with giant online subscription services like Amazon Prime for reading and initiate a digital footprint as this is more convenient for most of the customers who prefer doing everything online. This also cuts the costs of print media on Greyson and provides access to online archiving on the magazines. A digital subscription can still be used by their end client to serve advertising as well. Ads could be refreshed or rotated within the subscriber’s digital version of the magazine, increasing not only customer convenience but the convenience of stakeholders who depend on Greyson for print advertising.

As Greyson have a long-running strength of expanding their services by cross-selling, in addition to this we recommend focusing on the high-income people who are homeowners, ladies who are almost 83% of the population, young and teenaged kids at home to encourage these groups to have copies of magazines by giving away coupons, incentives on trips, local events for kids and restaurants if they have these magazine subscriptions.

Parents who teenage or school age kids are not as likely to renew their subscription. To combat this, Greyson could also create a kid friendly and teen friendly magazine that can be sent along with the parent’s subscription, for a nominal fee to encourage parents to renew.

We recommend Greyson to work closely with their data teams to target the identified groups from our model and find a lookalike audience and initiate a campaign for subscriptions or renewals by integrating to their networking accounts or shopping accounts. This can expand their customer based and allow them to use the given model to adjust their selling and renewal strategies.

**Reference**

[GREYSON Corp consulting](https://www.greyson.eu/index.php?lang=en)

<https://www.mindtools.com/pages/article/stp-model.htm>

<https://www.investopedia.com/terms/v/valueproposition.asp>