Domino's Pizza Churn Analysis

Desautels Faculty of Management, McGill University

MRKT 440: Marketing Analytics

Professor Ernan Haruvy

December 1, 2022

Alexandra Gorgeon Merino (261034270, Group 5)

Georgia Haydock (260903737, Group 5)

Benjamin Soroka (261028335, Group 4)

Jonathan Steinberg (260924571, Group 4)

Problem Statement

In 1960, the first Domino's pizza was created as a standalone restaurant in Ypsilanti, Michigan. Quickly franchised and then made public, the company saw a whirlwind of growth to arrive at its current status: a 12 billion dollar brand with close to 19,000 stores worldwide (Lock, 2022). Now, who does not love pizza? A surprising number of people when one looks at the monthly customer churn rate of the beloved pizza company, Domino's Pizza.

During February and March of 2016, Domino's experienced a significantly high month-to-month churn rate of 51.93%, in which 625,920 of 1,169,029 customers that made purchases in one month failed to make a purchase in the next. For any business, churn is exceptionally costly: On average, it is seven to eight times more expensive to acquire a new customer than to retain an old one (Voix, 2018). Additionally, according to a study done by Harvard Business Review, a 5% increase in customer retention (or a 5% decrease in customer churn) can increase profits by anywhere from 25-95% (HBR, 2000).

To limit its future losses, Domino's must address its high customer churn rate. Therefore, we will analyze several segments of customers and identify which group should be targeted with a coupon campaign to prevent them from churning. This leads us to our research question: which alternative group of consumers should Domino's allocate its coupon budget towards in order to most effectively reduce churn? For our analysis, *churn* is the number of customers who made a purchase (or multiple) in the previous month and failed to make a purchase in the current month. For example, if a customer makes at least one order in January but fails to do so in February, we consider that customer to have churned.

Through the creation of a churn prediction model and a Recency, Frequency, and Monetary (RFM) clustering analysis, we identified four groups of customers to potentially target:

all customers predicted to churn (those with over a 50% chance of churning), customers with a high risk of churning (those with over a 62% chance of churning), and two groups with undesirable levels of RFM (those who achieved significantly low RFM compound scores). We also analyzed a random subset of customers to be used as a benchmark for the four alternative groups.

We will use a collection of nine KPIs to decide which group is most effective to target. The first three are extracted directly from the RFM analysis: Recency, Frequency, and Monetary. Recency will account for the average days elapsed since the last purchase of alternative group X, Frequency details the average amount of orders made in the month by alternative group X, and Monetary is the average amount spent on orders in the month by this same alternative group. Average customer lifespan will also be used as a KPI, calculated by dividing the sum of customer lifespans (days between the first and last purchase in the month) by the number of customers in the group. While Recency is most desirable when low, Frequency, Monetary, and customer lifespan are ideally high.

Several KPIs that were included in the dataset will also be used in our analysis: coupon usage, discount amount, and profit. Coupon usage is the average amount of coupons used in a month, discount amount is the average amount of money taken off orders in the month, and profit is variable MenuAmount minus variable IdealFoodCost. The two discount-related KPIs will give us a sense of each group's willingness to use coupons, as we would ideally like to offer coupons to those willing to use them.

The last two main KPIs we will analyze are predicted churn rate and actual churn rate. Predicted churn rate refers to the average percent of customers predicted to churn in group X. The higher the predicted churn, the more targeting that group with coupons is justified before the start of the campaign and is, therefore, one of our most critical KPIs. Actual churn is the proportion of customers in group X that actually churned, and highlights how many of our coupons will go to customers who churned after all. While it will not be used to decide between alternatives since it can only be measured once March concludes, it can be used as a proxy for the effectiveness of the coupon campaign once we choose a group of customers to target.

Analysis

Original Data Description

The original dataset used for this analysis consists of 3,010,423 Dominos orders by 1,169,029 customers between January 1st, 2016, and March 31st, 2016, inclusive. Each observation represents an order and is accompanied by a unique identifier for each customer (AddressId) and store (StoreNum), as well as the date and time the order was placed (DateOfOrder and OrderTime). Every order also has various monetary amounts associated with it; the ones relevant to our analysis are the final amount that the customer pays (OrderAmount), the amount deducted due to coupons or other promotions (DiscountAmount), and the cost of the ingredients (IdealFoodCost). Other features that will prove to be useful include how the customer made the order (OrderType and OrderMethodDesc) and a unique code and description for each type of coupon (CouponCode and CouponDesc). Several features indicate the contents of the order (such as Drink20ozCount and DessertCount); while these are not used to directly analyze churn, we will analyze the contents of our chosen group's order to offer them coupons of food types they have never tried before.

Data Preparation and Methodology

Data preparation begins with creating two new features: Timestamp and Profit. The Timestamp feature refers to elapsed time in seconds since epoch time, which in Python is

January 1st, 1970, and is created by merging the DateOfOrder and the OrderTime features (n.d, Python). The profit feature is created by simply subtracting IdealFoodCost from OrderAmount. Next, the original dataset is separated into three distinct datasets, one for each month. The January dataset contains 608,213 unique customers with 975,395 (32.4%) of total orders, February contains 596,878 unique customers with 959,928 (31.87%) of total orders, and March has 644,277 unique customers with 1,075,100 (35.71%) of total orders.

To predict month-to-month churn, we create a logistic regression model. Specifically, we utilize data from the January dataset to predict churn in February and data from the February dataset to predict churn in March. In other words, we will compute the Recency, Frequency, and Monetary for each customer that made a purchase in January and use their values to predict if a customer will churn in February. Similarly, in February, the RFM values for each customer will be used to predict whether or not that customer will purchase in March. We use January RFM & February churn values as our training set and February RFM and March churn as our test sets.

Feature Engineering

Recency is calculated by subtracting the timestamp of a customer's most recent order in the current month by the exact timestamp that the current month ends and dividing the difference by 86 400, which corresponds to the number of seconds in a day. In other words, Recency represents the number of days between a customer's most recent purchase and the end of the month in question. Frequency and Monetary were computed by taking the sum of each customer's orders and OrderAmount during the current month, respectively. Finally, we derive churn by iterating through all AddressIds in the current month and checking if they appear in the next month's dataset. Other metrics such as customer lifespan, profit, coupon amount, and discount amount were also calculated for each customer during this process to be used as KPIs later in our analysis.

The February RFM values for each customer were also used in a clustering analysis to gain insight into which segments of customers are most valuable to Dominos. We chose to construct four clusters since finding the optimal number of clusters using the silhouette method was too time-consuming due to the large size of our dataset. To quantify customer value in terms of RFM, we computed an RFM compound score for each customer using the following formula:

$$\boxed{compoundscore_i = 0.33 * STD(29 - Recency_i) + 0.33 * STD(Frequency_i) + 0.33 * STD(Monetary_i)}$$

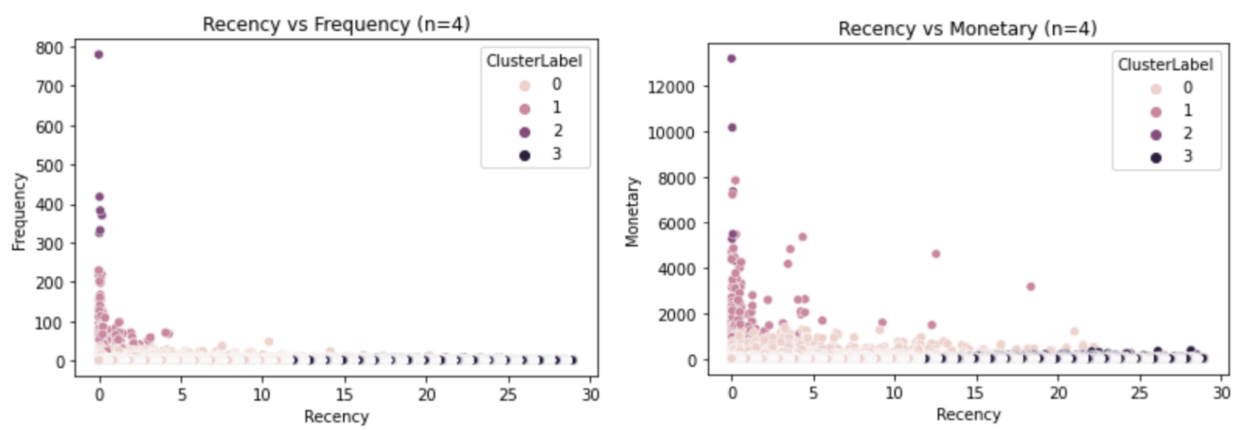
The compound score is a weighted average of a customer's standardized Recency, Frequency, and Monetary values; the higher the compound score, the more favorable the customer is believed to be. However, since high Recency values are not desirable, we modify it by subtracting it by 29, essentially taking its reciprocal since there are 29 days in February 2016. We then compute the average compound score of customers in each cluster to understand which clusters contain the lowest-performing customers.

Results

Logistic Regression and Clustering

Our logistic regression model achieves a 63.16% accuracy, meaning that we were able to correctly predict whether a customer would churn or not in March using February RFM data 63.16% of the time. Furthermore, our model achieves an exceptionally high recall of 80.85%, meaning that among all customers that truly churn in March, there is an 80.85% probability that our model identifies them correctly. The precision of our model is the lowest of these three metrics at 59.9%, indicating that for all customers predicted to churn, the probability that the model classifies them as such is 59.9%. The coefficients for Recency, Frequency, and Monetary were found to be 0.0111, -0.6666, and -0.0005, respectively. Notice that the magnitude of each of these coefficients makes sense; the customers most likely to churn are the ones with high Recency, low Frequency, and low Monetary.

The clustering analysis revealed two weak customer groups regarding their RFM compound scores. Clusters 1 and 2 achieve high respective compound scores of 19.7 and 105.1, while Clusters 0 and 3 have less desirable compound scores of 0.32 and -0.46, respectively. The inferiority of Clusters 0 and Cluster 3 can be visualized by analyzing the scatter plot below: Both Cluster 0 (represented by the lightest dots) and 3 (represented by the darkest dots) have low Monetary and Frequency values, while only Cluster 0 ranks low in terms of Recency.

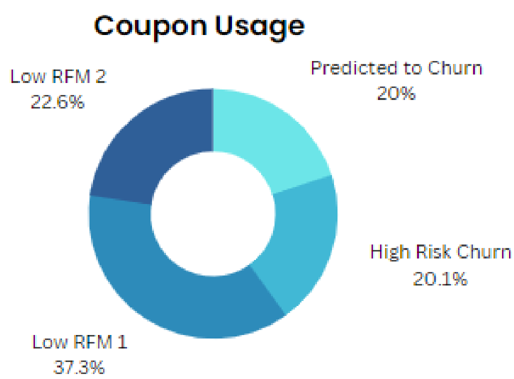


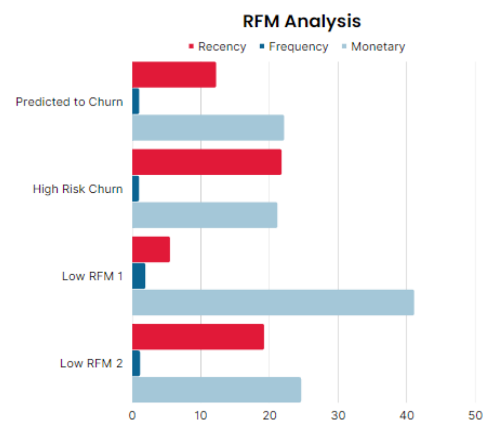
Alternatives Comparison

We will now compare the four alternative groups of customers through our nine KPIs to determine which is best to target with coupons. An exhaustive view of the values of all KPIs with respect to each alternative group of customers can be found in Table 1 in the Appendix.

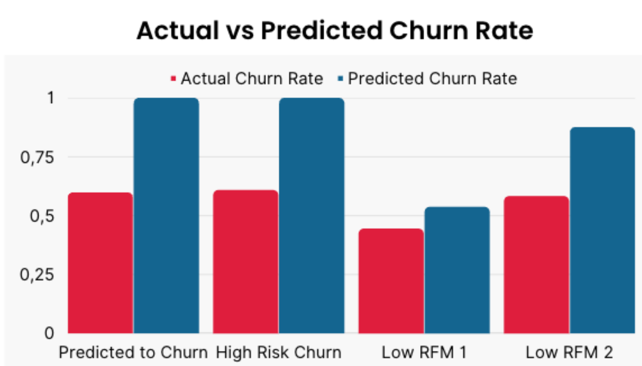
The customers in RFM Group #1 have many favorable characteristics that incline us to retain them. Firstly, they have the highest coupon usage and discount amount out of all groups, making them a significant receptor to coupons should we offer them. They also rank the best in terms of Recency, Frequency, Monetary, customer lifespan, and profit, the latter of which is over

$10 higher than the next highest group. Nevertheless, one may wonder if targeting this group is worth it because they have the lowest predicted churn rate (53.73%), implying we would unnecessarily be giving out coupons to 46.28% of customers according to our prediction model.





The customers in RFM Group #2 have less favorable KPI values compared to those in RFM Group #1, however they still have some attractive aspects: They have the second highest Recency, coupon usage, and predicted churn rate. However, their values for the rest of the KPIs are less significant than other segments, complicating its differentiation. Interestingly, RFM Group #2 has a high predicted churn rate of 87.59%, which is much higher than that of RFM Group #1 and our random subset, which had a 67.86% predicted churn rate. Although a predicted churn rate of 87.59% is favorable, it is not as high as our two predicted churn groups, who naturally have a predicted churn rate of 100% each.



The group consisting of all customers predicted to churn have poor RFM values and the lowest lifespan. Their coupon usage is also the lowest out of all groups, potentially making our campaign uninteresting to them. However, these weak values are somewhat expected; of course customers predicted to churn will not be the most engaging ones, and that is precisely what we are trying to fix through a coupon campaign. Their predicted churn rate is also 100%, as anyone with a churn probability of over 50% is considered churn in our model.

The high risk churn group are those with over a 62% chance of churning, which corresponds to the 75th percentile of all predicted churn probabilities in the month of February. This group performs the worst across our RFM KPIs, having the highest Recency, lowest Frequency, and lowest Monetary compared to any other group. An advantage of the high risk churn group is its size: they have less than half the amount of customers than the group consisting of all customers predicted to churn, allowing Domino’s to target a more curated group of customers that fits better in their budgetary constraints.

Recommendation

Given the performance of our four alternatives on our nine KPIs, we have decided to target the high risk churn segment as we want to focus our efforts on a niche group of customers with the highest predicted churn rate.

Place

Out of the 149,220 customers in this group, 76,391 (51.2%) ordered from the website, 70,803 (47.4%) ordered from the phone, and 2,026 (1.4%) ordered by simply walking into a Domino's location. Those who ordered through the website would receive coupons through email, bearing no cost to Dominos. On the other hand, those who ordered via the phone would be sent coupons through the mail, assuming their email addresses were recorded when they placed the order. With this strategy, there would be a small mailing cost incurred by Domino's. Unfortunately, for the 2,026 people who ordered when they walked into the Domino's location, we assume there was no personal information taken that could be used to send them coupons, shrinking our targeted group size to 147,194.

Product

Ideally, these high risk churn customers would be given coupons for products they have never tried before, assuming their distaste for previously bought meals contributed to them churning. Three of the least purchased products in February by high risk churn customers were chicken wings, desserts, and drinks, so the suggested coupons should include these products. The most popular coupons including wings among all customers in February was the LG 2 TOP and 14 wings at $19.99 (354 orders), and is suggested for those who have never ordered the chicken wings. The most popular coupon including a dessert was the 2L and Choice of Cinna Stix or Lava Cakes at $5.99 (253 orders), and is suggested for those who have never ordered a dessert. The second most popular coupon including a drink was the LG 1 TOP Pizza, 10pc chicken, and 2-20oz drinks at $19.99 (68 orders), and is suggested for those who have never ordered a dessert. Many of these coupons contain other food items in addition to the ones that are less frequently purchased, which exposes customers to more food types that they potentially will enjoy. In addition, Dominos can mandate that these coupons must be used in separate orders to generate higher frequency.

Price

To determine the price of this promotion campaign, we added each coupon price together and multiplied the result by the sum of customers who ordered on the website or by phone. The total price of the campaign ends up being $6,766,508.18, which is below the budget of 10 million dollars by more than 3 million. Therefore, Dominos can run trials to see which coupons are most effective and make changes if needed with the remaining budget.

Appendix

	Predicted to Churn	High Risk Churn	Lowest RFM #1	Lowest RFM #2	Random Subset
NumberOfCustomers	405034	149220	347267	249421	287735
Recency	13.20	21.74	5.49	19.19	11.21
Frequency	1.00	1.00	1.90	1.14	1.61
Monetary	22.10	21.12	41.05	24.58	34.83
Lifespan	0.00	0.00	6.42	0.71	4.04
Profit	16.42	15.70	30.18	18.22	25.62
CouponsUsage	0.64	0.64	1.19	0.72	1.01
DiscountAmount	5.45	5.30	10.09	6.06	8.57
ActualChrunRate	0.60	0.61	0.45	0.58	0.50
PredictedChurnRate	1.00	1.00	0.54	0.88	0.68

Table 1: All KPI Values for Each Alternative Group

Works Cited

The economics of E-loyalty. (2000). Retrieved November 30, 2022, from https://hbswk.hbs.edu/archive/the-economics-of-e-loyalty

Voix. (2018, August 25). Why should a restaurant invest in customer retention? Retrieved November 30, 2022, from https://medium.com/@voixai/why-should-a-restaurant-invest-in-customer-retention-9bad85b8d40b

How to get Timestamp in Python? | Flexiple Tutorials | Python. (n.d.). https://flexiple.com/python/python-timestamp/