Offer Effectiveness Analysis for Star Bucks

Using Starbucks's simulated customer data

Project Overview

Starbucks is a well-known coffee shop with locations all over the world. Starbucks, like most other stores, has a mobile app via which users can receive special discounts from time to time. An offer might be as simple as a drink commercial or as complex as a discount or a BOGO deal (buy one get one free). Every offer has a time limit before it expires, and some, such as discounts and BOGO, require a minimum purchase to be eligible.

Problem Statement

The most important question to address is which customer groupings respond well to which offer type. As a result, this challenge might be framed as a classification problem, with the goal of determining which types of people respond best to which types of offers.

High Level Methodology

Since we've created a decent comprehension of the business and its requirements, we'll go on solving the problem as below. Code base could be found in this ***github*** repository.

1. Data Understanding
2. Data Preparation
3. Explonatory Data Analysis (EDA)
4. Modeling & Evaluation
5. Discussion & Conclusion

1. Data Understanding

It is vital to understand the data in order to design the startergy to answer the question with exploratory analysis and modelling .The raw data is contained in 3 files.

Portfolio.json

This contains offer ids and metadata about each offer (duration, type, etc.).

- id (string)—offer id
- offer_type (string)—type of offer i.e: BOGO, discount, informational
- difficulty (int)—minimum required spend to complete an offer
- reward (int)—reward given for completing an offer
- duration (int)—time for offer to be open, in days
- channels (list of strings)

BOGO deals: These are the deals where you buy one and receive one free. Validity will be 5 or 7 days, with two sorts of schemes offering various rewards for each period. All four schemes will be distributed via cellphone and email, with some also being disseminated via web or social media. What's interesting is that as the validity period gets shorter, more communication channels become involved. Because it's a buy one, get one free deal, the return for spending is 1:1.

Discounts: These are for a longer validity term than bogo, which is either 7 or 10 days, and all schemes are communicated via online and email, with one scheme being highly communicated over four channels for each validity period.

Informational offers: There are two informational schemes with no offers (no benefit to the customer) that are communicated via email, smartphone, or online / social media.

Profile.json

This contains demographic data for each customer.

- age (int)—age of the customer
- became_member_on (int)—date when customer created an app account
- gender (str)—gender of the customer (note some entries contain 'O' for other rather than M or F)
- id (str)—customer id
- income (float)—customer's income

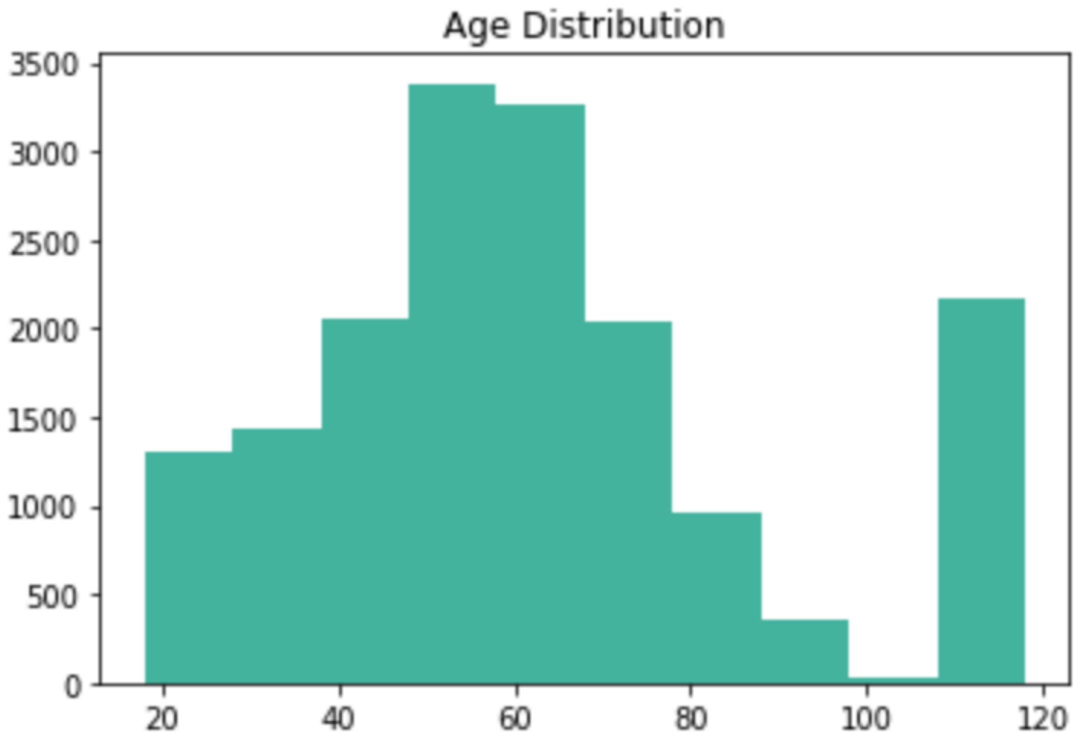


Image by Author: Age distribution with outlier age of 118

This dataset contains 17K customer records, with roughly 13% of demographic data missing. Age appears to have outliers and when analysed it showed that set of customers have an age of 118. Gender and income values are missing and they appeared to be the same records which had age outliers. As a result, these records will be removed from further analysis since 13 percent of records had missing/outlier values for three columns.

Transcript.json

This dataset contains the customer's transaction history with a time stamp. The value of the transaction, the offer that was received, the time it was viewed, and the time it was completed are all listed here.

2. Data Preparation

This is one of the most crucial steps in the procedure. Outliers and missing values were mostly dealt with, on the profile dataset, as previously stated. Categorical variables will be one-hot-encoded in order to prepare the data for modeling. The data frames were then cleaned and merged to form a single master table at the customer level with few features developed and successes flags were also calculated. In the 2.2 section this would be further described. The final master table will have the below columns:

```
['gender', 'age', 'cust_id', 'became_member_on', 'income',
  'years_member', 'year_joined', 'gender_F', 'gender_M', 'gender_O',
  'frequency', 'recency', 'monetary_value', 'sucess_flag_bogo',
 'sucess_flag_discount', 'sucess_flag_informational', 'tot_success',
 'event_count_bogo', 'event_count_discount',
'event_count_informational','tot_count_offers_received']
```

2.1 Feature Engineering

Feature engineering is a step that we can take before we start modeling. The following features were developed based on the hypothesis below. It should be noted that even though the data was available on the channels used to communicate the offers, they were not used to create features given the fact that most channels were common across offers as described in the data understanding phase.

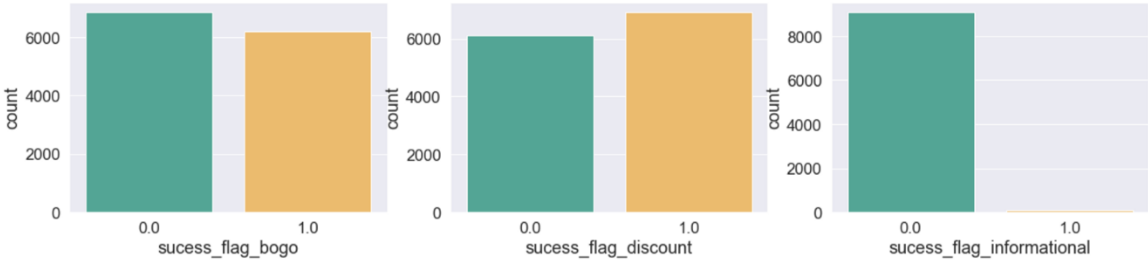
- years_member : Longer a customer with the company more likely to respond to an offer.
- recency: Customers who purchased more recently are more likely to purchase again than are customers who purchased further in the past.
- frequency: Customers who have made more purchases in the past are more likely to respond than are those who have made fewer purchases.
- monetary : Customers who have spent more in the past are more likely to respond than those who have spent less.

2.2 Success flag calculation

The success of each offer type, as well as the total number of offers, was calculated. If a customer viewed an offer before ingesting it, he or she was considered a success for discounts and BOGO. For informational offers, it is determined whether they completed a transaction after receiving the advertisement within the validity period provided in the portfolio data. Starbucks determines the validity period for informational offers based on their subject expertise and experience that a client will be influenced by an advertisement for this number of days. There were few instances where customer has consumed two offers in one go and if they have viewed the offer before the offers were taken as success too.

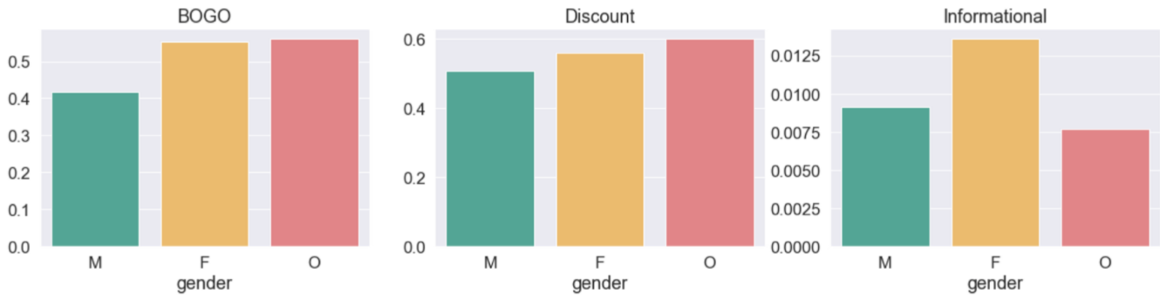
3. Exploratory Data Analysis

EDA is a phase where we can explore some interesting insights which will directly influence the modelling phase.



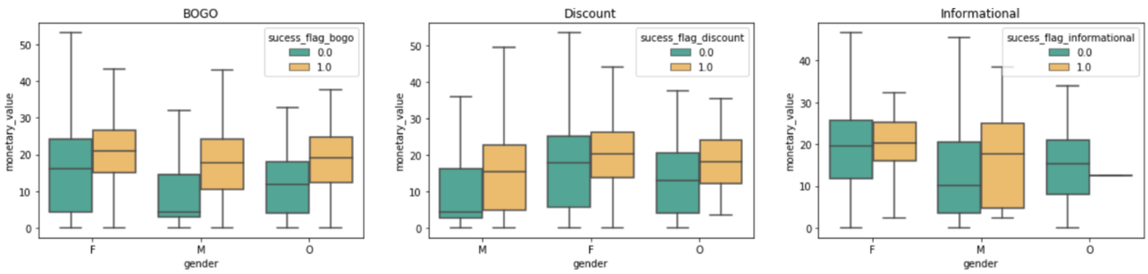
For each offer type success count plots

It was clear that informational offers had a very low success rate, whereas BOGO and discount offers had just a small difference on the number of successes and failures. Further it was very interesting to see regardless of offer type Females were more attracted to the offers.



Gender wise succes rates

The notebook can be used to access all of the analyses because there are so many to be stated here. In high-level, age does not appear to have a direct 1–1 relationship with success rates, but frequent, high-spending, or recently shopped clients appeared to have a higher success rate.



Gender wise distribution of spend distribution of customers with distinction of success rates

4. Modelling and Evaluation

As stated in the project overview, we will utilise classification models to identify successful offers, and the attributes that become relevant for classification will be significant features to identify which types of clients respond to each offer type.

It was noted during the EDA phase, the success rate of informational offers is quite low (only around 1% of total informational offers received clients), therefore we'll leave that out of the scope of the project until we have additional data to analyze. We'll also train two different models to recognize the success of BOGO and discount offers, so the features will clearly communicate the results for those specific offer types.

The data will be trained across four different types of base models, including decision tree, random forest, gradient bossing, and ada boost models, and then tuned to produce better outcomes, if required.

Metrics

The accuracy and the F Score will be the most important factors to consider while analyzing the findings. Accuracy indicates how accurate the result is, but the F score, which is the harmonic mean of precision and recall, balances measures such as false positive and false negative rates, which are also relevant in this topic. For example, if a success is mistakenly classified as a false negative, we may not make an offer to that customer, affecting the targeted results of making offers to more successful consumers in the long run.

We train on 80% of the data and test the findings on the remaining 20% test data to validate the results. The data will be scaled before training because the features are on various scales.

4.1 BOGO Discounts—Modelling

Target variable will be the flag whether the offer BOGO was successful or not and it will be trained across the features:
[‘age’, ‘income’, ‘years_member’, ‘gender_F’, ‘gender_M’, ‘gender_O’, ‘frequency’, ‘recency’, ‘monetary_value’]

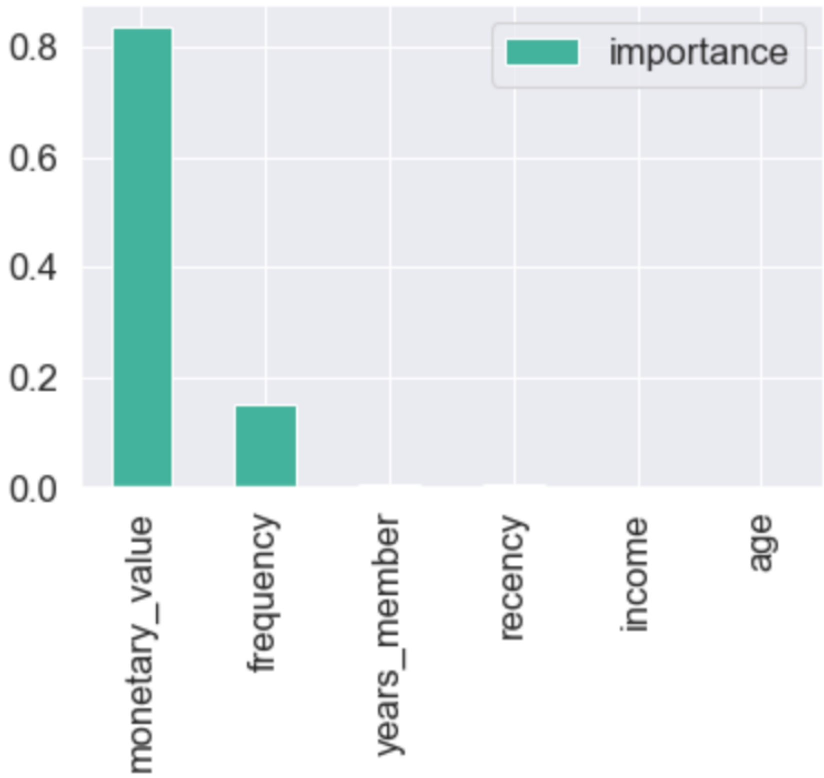
	DecisionTreeClassifier	RandomForestClassifier	GradientBoostingClassifier	AdaBoostClassifier
dataframe_name	[sucess_flag_bogo]	[sucess_flag_bogo]	[sucess_flag_bogo]	[sucess_flag_bogo]
training_score	0.999916	0.999916	0.708347	0.698904
testing_score	0.609106	0.674874	0.690051	0.692074
training_fscore_class_0	0.999928	0.999928	0.736095	0.727384
testing_fscore_class_0	0.663179	0.716304	0.71627	0.715842
training_fscore_class_1	0.999899	0.999899	0.674079	0.663779
testing_fscore_class_1	0.534351	0.619273	0.658491	0.663968

Results across different classifiers for BOGO offer customer data

If we train the data across several models without hyper parameter adjustment, as shown in the above image, the data will be overfitted in most situations, as training accuracy is higher than test accuracy. We’ll opt with gradient boost because the train test results are closer and there doesn’t seem to be much overfitting.

We can get a model with about 72% accuracy and F score in both train and test sets by hyper parameter adjustment (optimal parameters were found to be ‘learning rate’: 0.01,’max depth’: 3,’n estimators’: 100).

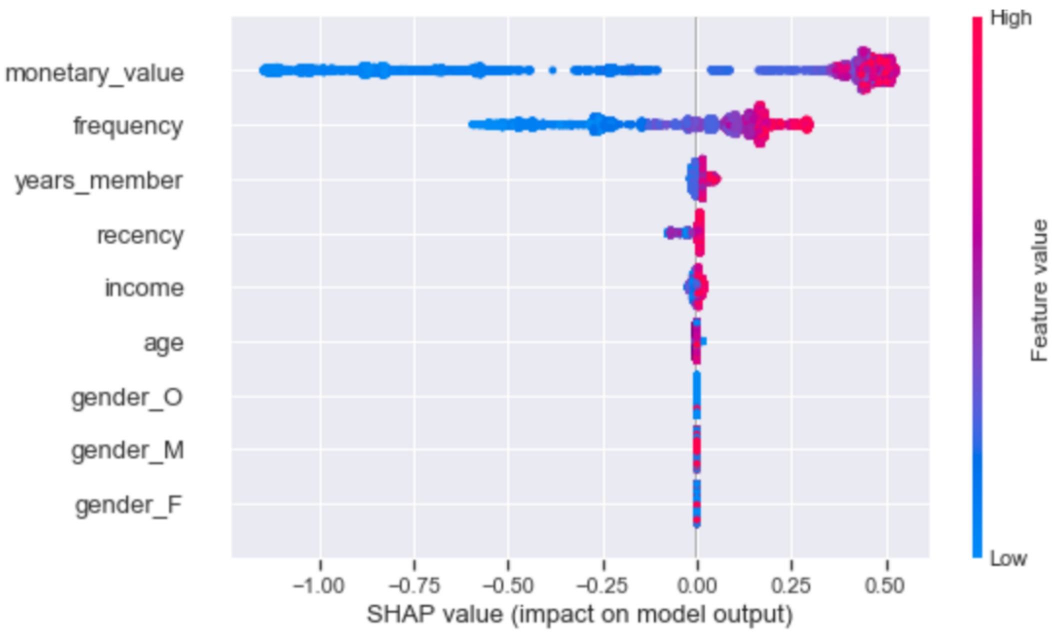
4.2 BOGO Offers—Feature Importance



BOGO offers: Gradient Boosting Feature importance

The success rate of BOGO offers seems to be affected by RFM features, membership period, income, and age. Customers' spending and frequency of visits appear to be the most essential factors among them.

Since the model is a black box model we can use the shap values to further understand the model.



It appears that the bigger the amount spent, the frequency with which the customer shops, or the number of years being a member, the better chance of a offer being a success.

4.3 Discount Offers—Modelling

Target variable will be the flag whether the offer Discount was successful or not and it will be trained across the features:
[‘age’, ‘income’, ‘years_member’, ‘gender_F’, ‘gender_M’, ‘gender_O’, ‘frequency’, ‘recency’, ‘monetary_value’]

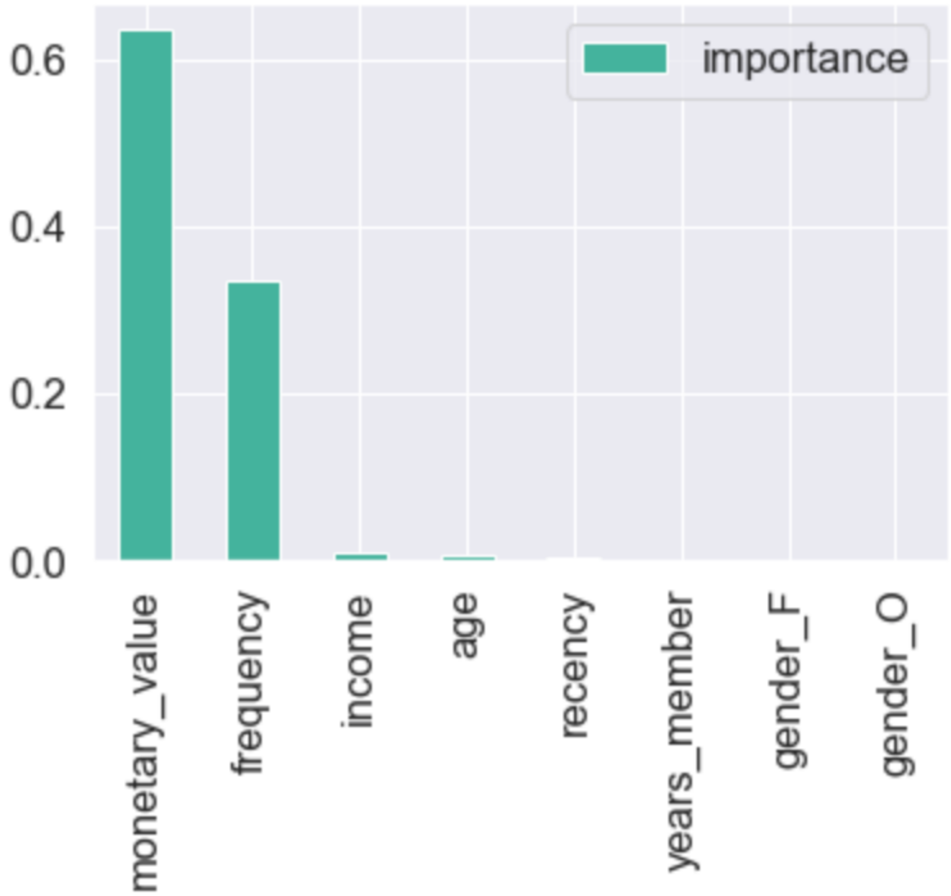
	DecisionTreeClassifier	RandomForestClassifier	GradientBoostingClassifier	AdaBoostClassifier
dataframe_name	[sucess_flag_discount]	[sucess_flag_discount]	[sucess_flag_discount]	[sucess_flag_discount]
training_score	1.0	0.999904	0.706604	0.693185
testing_score	0.597164	0.66079	0.68992	0.68417
training_fscore_class_0	1.0	0.999897	0.635291	0.635629
testing_fscore_class_0	0.569791	0.601531	0.617494	0.627823
training_fscore_class_1	1.0	0.99991	0.75459	0.735038
testing_fscore_class_1	0.621261	0.704705	0.739285	0.725699

Results across different classifiers for Discount offer customer data

Without pruning, all of the models (especially the decision tree model) appear to be overfit to the data, with the training score being higher and the test score being very low.

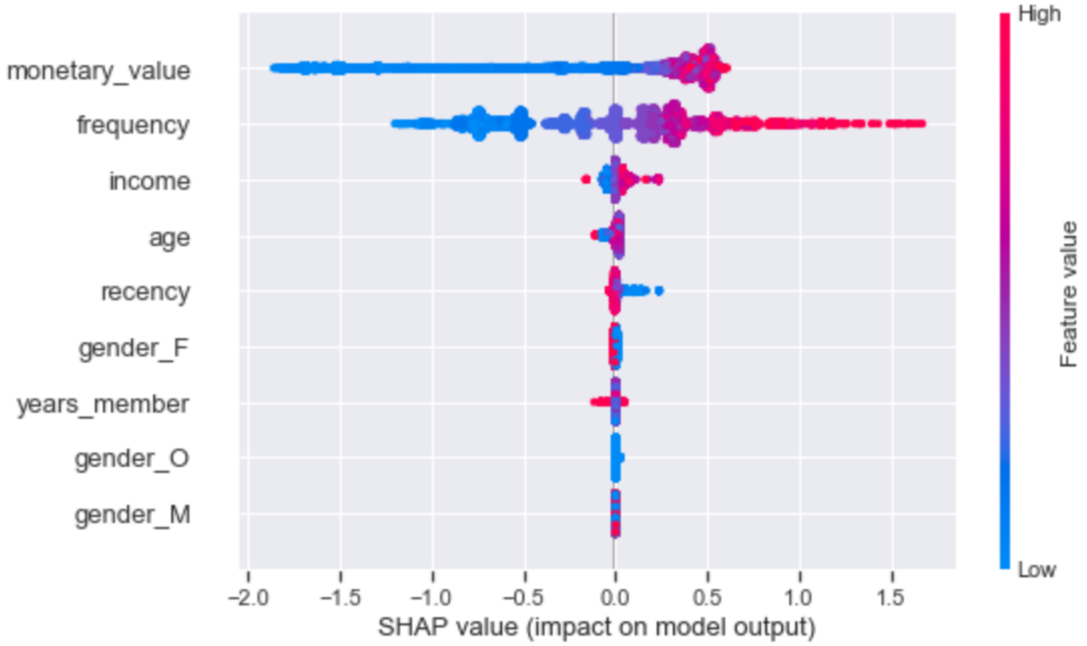
The gradient boost model was decided to go on with because the train test results are closer and there is not much evidence of overfitting. The optimum parameters for the model were ‘learning rate’: 0.01,’max depth’: 3, and ‘n estimators’: 250 after hyper parameter adjustment. Train and test sets were closer to 69 percent accurate, with train and test Fscores of 71 percent and 61 percent, respectively. More pruning of the model could improve these even more.

4.3 Discount offers—Feature Importance



Discount offers: Gradient Boosting Feature importance

The success rate of Discount offers seems to be affected by RFM features, income, age, membership period and age. Customers' spending and frequency of visits appear to be the most essential factors among them, similar to BOGO disocunts.



It appears that the bigger the amount spent, the frequency with which the customer shops, or the customers with higher income has the higher chance of offer being a success customer.

5. Discussion and Conclusion

Reflection

This was a full-fledged data science study that began with a business understanding and ended with technical advice. It was discovered that high-spending, high-frequent customers who have been with Starbucks for a long time were more successful in BOGO discounts, whereas high-spending, high-frequent customers with a high income range appear to be more effective in discount offers.

Improvements

We covered data cleansing, modeling, and interpretation in this project, and there are just a few enhancements that might be made. More features could have been developed, RFM features could have been grouped together with weights and created RFM clusters using domain knowledge, more model pruning could have occurred, and finally a FLASK web application for Starbucks could have been developed to query usergroups for a specific offers, making the implementation easier.