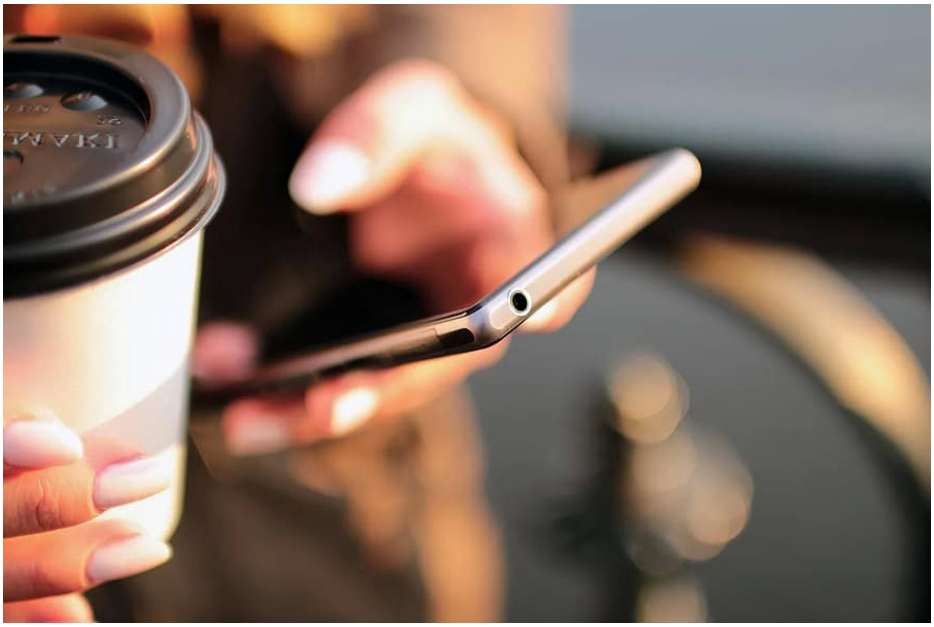
To offer or not to offer? A Starbucks Machine Learning Project



Looking for those free coffee rewards. Source Pikist.

I'll be the first to admit, I am an avid coffee drinker. But, I'm not committed to a single chain. On my phone, I have apps for the Coffee Bean, Starbucks, and Dunkin all downloaded. Most of the time, proximity will dictate which store I visit. However, increasingly as I gain more and more loyalty points with each chain, I am driven to go to a particular store because of a special offer notification I received.

How does Starbucks know when to send me those rewards? How do they choose between a buy one get one free offer, or a $2 off your next order offer, or simply a bonus star reward?

This was the question I sought to answer in my capstone project for the Udacity Data Science Nanodegree.

If you are interested in following along with the code, view the project on Github.

. . .

TL, DR:

- Females spend more money on average than males at Starbucks.
- Users who do not completely fill out their membership profile typically spend less, suggesting that they also would not benefit from “craft drink” related offers (i.e. try our new venti magic unicorn frappuccino with extra sprinkles and pixie dust and get 20 extra stars!).
- Offers where a discount is provided are more likely to be completed than buy-one-get-one offers.
- To get maximum completion, include social media in your distribution plan. The most successful and second-to-least successful offers were almost identical in all ways except that the most successful one was spread via social media and the other wasn’t.
- With machine learning, I validated that social media was an influential factor in whether an offer would be completed.
- Other important factors included offer duration, user age, and how much the reward was for. Unimportant factors included email, how long the user had been a member, and their gender.

. . .

The Data

To complete the project, Starbucks provided Udacity with 3 key datasets:

- portfolio: data about all 10 offers provided in the sample data
- profile: data about all the users who had transactions or offer notifications
- transcript: data about all transactions that occurred, be they purchases or interactions with offers

Diving into the nuances here, there are a few more things to note about each dataset. Portfolio specifies which channels a specific offer was released through, including email, mobile, social, or via the web. There is also a difficulty score, the dollar amount that must be spent for the offer to be completed. The reward amount, duration before expiry, and offer type are also specified.

Profile includes the user's gender, income, age, and date they became a member. Some of the demographic data is missing, however.

Transcript includes the event recorded and either the transaction amount, or whether an offer was received, viewed, or completed.

. . .

Problem Statement

The goal with this data is to identify trends in purchases and offer completion. Ideally, with this project, I will be able to predict whether a user will come into the store to purchase something because they received an offer.

There are a few considerations to keep in mind, therefore. If someone was going to come in and buy $10 of coffee anyways, offer or not, we don't necessarily want to send them a $2 off a $10 purchase offer.

Additionally, it may be the case that users will complete an offer without even reading it and knowing it exists. Again, in these cases, the offer has not met its goal of changing behavior.

Therefore, when parsing the data, I made sure to only consider an offer completed if the user viewed the offer and completed it before the offer expired.

. . .

Data Cleaning

Luckily, there was not much cleaning to do with these datasets. However, a few changes were made for ease of analysis.

First, the offer id's were a long, seemingly random string. These were replaced with simple 2 character codes. The first letter denotes whether the offer is informational, a discount, or buy-one-get-one (I, D, and B) and a sequential number.

Next, columns like gender and offer distribution method were remapped using dummy variables.

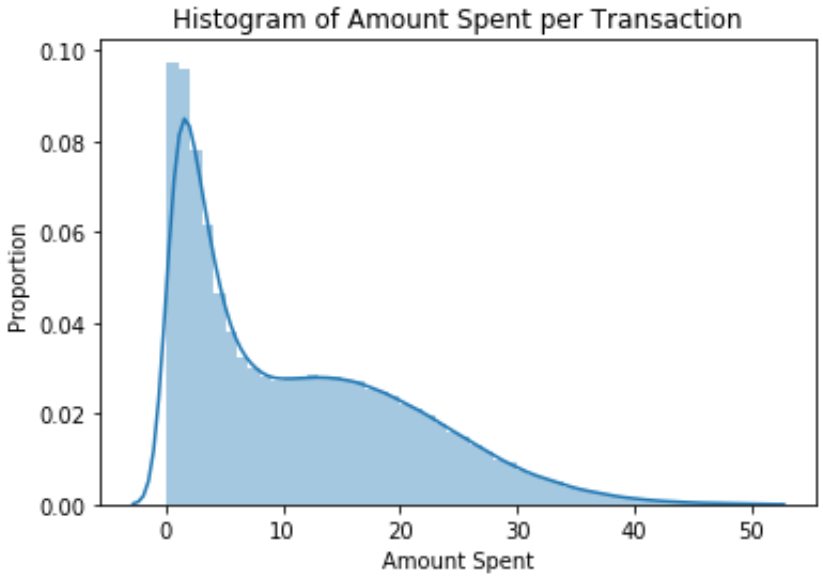
Data was also checked for completeness and correctness. It was found that a significant number of ages were inputed as 118, a clear error. These values were replaced with NaN.

. . .

Data Exploration

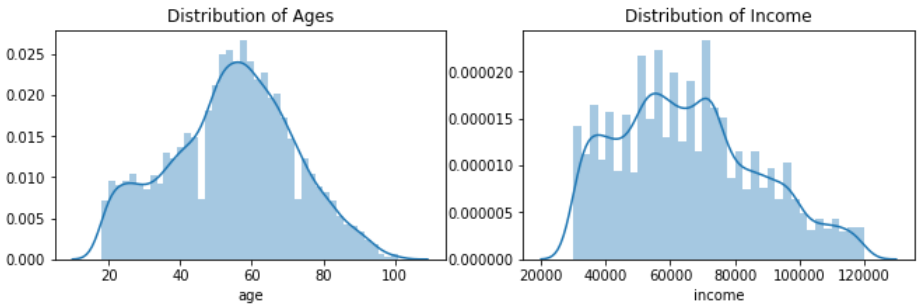
Before we get into the machine learning (which is what you are all here for, I know), I started with preliminary data exploration. This helps identify machine learning questions, validate conclusions drawn from machine learning, and find basic statistical descriptors of the dataset that cannot be identified through machine learning alone.

First, I explored the amount spent per transaction.



As expected for anyone who frequents a coffee shop, the vast majority of transactions are below $8. It is interesting that there is a large peak around $1–2, perhaps these are those people who just go in and get a cup of Pike Place, with room.

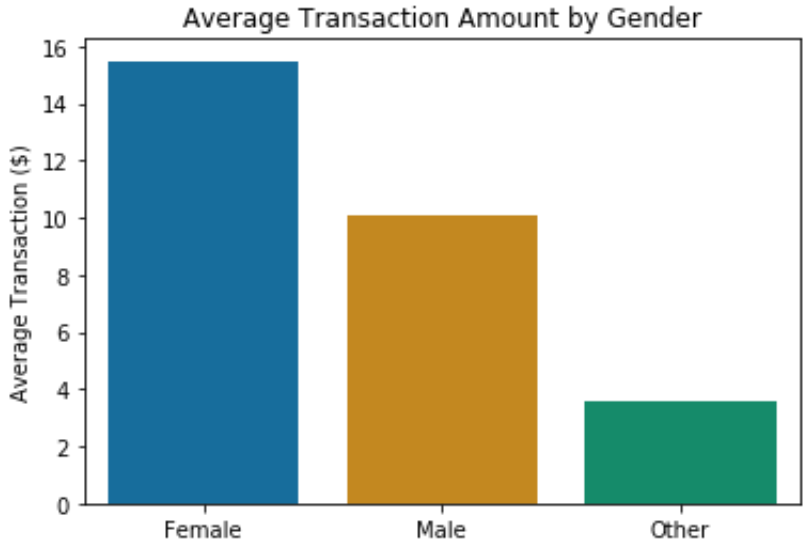
Next, I was interested in user demographic data. I only had age and income to consider.



This looks like a pretty standard user base, though I would have expected a lower peak for age distribution.

Also, shoutout to the 100 year old grammy at Starbucks.

Now, anyone who has been to a Starbucks knows the meme of the teenage girl with a frappuccino. So, I also wanted to look at average expenditure by gender. There was some “other” data for gender, either signifying that gender data was not collected for that user, or that the user identifies as a gender outside the binary. It would have been nice to differentiate between those two, but alas.



Interesting. Females typically spend around $5 more than males. Since each craft drink is around $5, this suggests that females are buying multiple drinks more often than males (potentially buying for the family or group of friends).

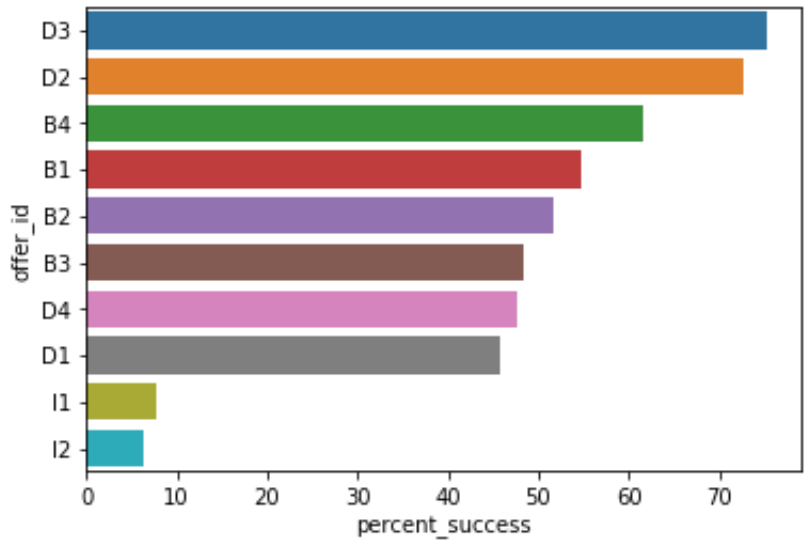
This graph also gives us some insight into the “other” category. It may be that these users are not fully registered with the membership and are those people only coming in to buy their cup of black coffee, and not particularly interested in special offers for craft drinks.

Of course, this particular insight will require more data to validate.

Next, I was interested in which offers were most likely to be completed. As a reminder, the codes here refer to whether the offer is a straight discount (i.e. $2 off a $10 order) or a buy-one-get-one offer. There are also informational offers, but these simply inform users of a new product, for example, and are not actionable. Each offer was also distributed in any of 4 ways: email, mobile, social media, or the web.

	offer_id	count	percent_success	reward	difficulty	duration	email	mobile	social	web
6	D3	6652	75.285628	2	10	10	1	1	1	1
5	D2	6655	72.742299	3	7	7	1	1	1	1
3	B4	6576	61.618005	5	5	5	1	1	1	1
0	B1	6683	54.646117	10	10	7	1	1	1	0
1	B2	6593	51.721523	10	10	5	1	1	1	1
2	B3	6685	48.287210	5	5	7	1	1	0	1
7	D4	6631	47.730357	2	10	7	1	1	0	1
4	D1	6726	45.762712	5	20	10	1	0	0	1
8	I1	6657	7.721196	0	0	4	1	1	0	1
9	I2	6643	6.277284	0	0	3	1	1	1	0

First off, it is important to notice that each offer was distributed roughly the same number of times, so there is no bias there. The last two rows can also largely be ignored because they are informational offers, not discounts or BOGOs.



Some key takeaways from this analysis are that discounts are more likely to be completed than BOGO offers. Additionally, all of the top performing offers were distributed in all ways possible, but specifically social media. When only social media was removed, those offers went to the bottom of the ranking, even though they were also discounts.

In fact, lets compare D3 and D4. D3 was the best performing offer, but it is almost identical to D4, expect for one key element: social media. D3 also lasted a few days longer, but considering that the worst performing offers all were not on social media is telling.

Machine Learning

Now for the fun part, building a machine learning model to predict which factors increase offer completion.

The response variable was whether the offer was successful, coded as a 0 for unsuccessful and 1 for successful. An offer was defined as successful if it was read by the user and fulfilled before it expired.

These are the factors included in the machine learning algorithm, and the values for the first 4 transactions considered. Note that start_year and start_month refer to when the user became a member.

	time	total_amount	reward	difficulty	duration	web	email	social	mobile	bogo	informational	discount	gender	age	income	start_year	start_month
0	0.0	37.67	5	5	7	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0	75.0	100000.0	2017	5
1	7.0	49.39	0	0	3	0.0	1.0	1.0	1.0	0.0	1.0	0.0	0	75.0	100000.0	2017	5
2	17.0	48.28	10	10	7	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0	75.0	100000.0	2017	5
3	21.0	48.28	5	5	5	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0	75.0	100000.0	2017	5

A 25% training split was completed to set up for the machine learning pipeline.

With this, I was ready to implement three different methods:

- Decision Tree Classifier
- Logistic Regression
- Random Forest Classifier

And here are the related scores for each model:

```
              precision    recall  f1-score   support

           0       0.88      0.89      0.88      8723
           1       0.87      0.86      0.87      7903

    accuracy                           0.87     16626
   macro avg       0.87      0.87      0.87     16626
weighted avg       0.87      0.87      0.87     16626

Decision Tree: 0.8727820046192608
              precision    recall  f1-score   support

           0       0.78      0.88      0.83      8723
           1       0.85      0.72      0.78      7903

    accuracy                           0.81     16626
   macro avg       0.82      0.80      0.81     16626
weighted avg       0.81      0.81      0.81     16626

Logistic Regression: 0.8044773113018622
```

```
Auc score: 0.9165310505899159
Precision: 0.8945388349514564
Recall: 0.9326837909654562
```

Random Forest Classifier

Using each model, I was also able to determine which explanatory factors were most impactful on whether an offer would be completed:

Weight	Feature
0.4076 ± 0.0053	total_amount
0.1335 ± 0.0012	duration
0.0845 ± 0.0016	income
0.0684 ± 0.0017	age
0.0637 ± 0.0017	time
0.0551 ± 0.0017	social
0.0455 ± 0.0015	start_year
0.0407 ± 0.0009	start_month
0.0380 ± 0.0013	reward
0.0192 ± 0.0007	gender
0.0153 ± 0.0006	mobile
0.0147 ± 0.0008	difficulty
0.0090 ± 0.0003	web
0.0063 ± 0.0005	bogo
0.0045 ± 0.0004	discount
0 ± 0.0000	informational
0 ± 0.0000	email

Decision Tree

Weight	Feature
0.2902 ± 0.0028	total_amount
0.0069 ± 0.0011	reward
0.0025 ± 0.0010	duration
0.0011 ± 0.0018	difficulty
0.0011 ± 0.0004	age
0.0008 ± 0.0004	social
0.0008 ± 0.0002	informational
0.0003 ± 0.0001	bogo
0.0003 ± 0.0001	web
0.0003 ± 0.0001	mobile
0.0001 ± 0.0001	discount
0.0001 ± 0.0000	gender
0.0000 ± 0.0000	start_year
0 ± 0.0000	email
-0.0001 ± 0.0001	start_month
-0.0001 ± 0.0006	time
-0.0006 ± 0.0005	income

Logistic Regression

Weight	Feature
0.3179 ± 0.0041	total_amount
0.0090 ± 0.0009	social
0.0057 ± 0.0004	reward
0.0057 ± 0.0006	income
0.0045 ± 0.0007	difficulty
0.0041 ± 0.0003	start_year
0.0031 ± 0.0004	time
0.0028 ± 0.0002	mobile
0.0027 ± 0.0004	age
0.0022 ± 0.0003	discount
0.0015 ± 0.0003	web
0.0014 ± 0.0007	gender
0.0014 ± 0.0002	start_month
0.0011 ± 0.0002	bogo
0 ± 0.0000	email
-0.0007 ± 0.0011	duration
-0.0010 ± 0.0004	informational

Random Forest

Each model gave a slightly different result, but overall it appears that important factors include offer duration, user age, and how much the

reward was for. Unimportant factors included email, how long the user had been a member, and their gender. The Random Forest model also identified social media distribution as an important factor, which I was able to identify visually earlier.

. . .

Conclusions

Through this project, I was able to build a machine learning model and explore the data to find some key traits that would make an offer more likely to succeed.

Some improvements to the dataset would have helped refine the predictions I was able to make. For example it would have been great to have location data or longer user history data. Specifically, it would have been interesting to see if users who always order the same few things were swayed to purchase new menu items by offers. This was one significant drawback of this dataset as it does not specify what exactly was purchased.

I'm interested to hear your thoughts! Have you ever seen an offer on your Starbucks app (or comparable for another store) that seemed to be tailor made for you? With more and more data available for users, it seems like personalized offers are only going to increase in frequency.