Starbucks Python Data Analysis Project

Debbie Trinh

March 3, 2023

Introduction

Starbucks is a Business To Consumer Company that is the largest coffeehouse chain in the world. It offers specialty coffee, food items, handcrafted beverages, and merchandise.

The goal of this project is to understand the marketing attribution of the Starbucks rewards offer marketing strategy which drives user engagement and transactions.

Agenda

Understand the Customer: Descriptive Statistics

Number of Members / Aggregate Spend

Spending by Age Group using Sturges Rule

Spend Range / Number of Members Spent vs Not Spent

User Tenure Distribution

Influence of Reward Offers on User Behavior

Actions taken in Marketing Funnel

Customer Personas that reward offers appeal to

Effective Offer Types Variants / Gender that completes most offers

Channel Effectiveness

Reward Offer Response Rate

Recap of Insights / Marketing Strategy Actionable Recommendation

Next Steps to Improve Marketing Strategy via ML

Understand the Customer: Descriptive Statistics

Understand the Customer:

How many Starbucks members are in this dataset?

In aggregate, how much have these members spent?

```
1 print("There are",len(pd.unique(transcript['person'])),"unique Starbucks members in this dataset.")
```

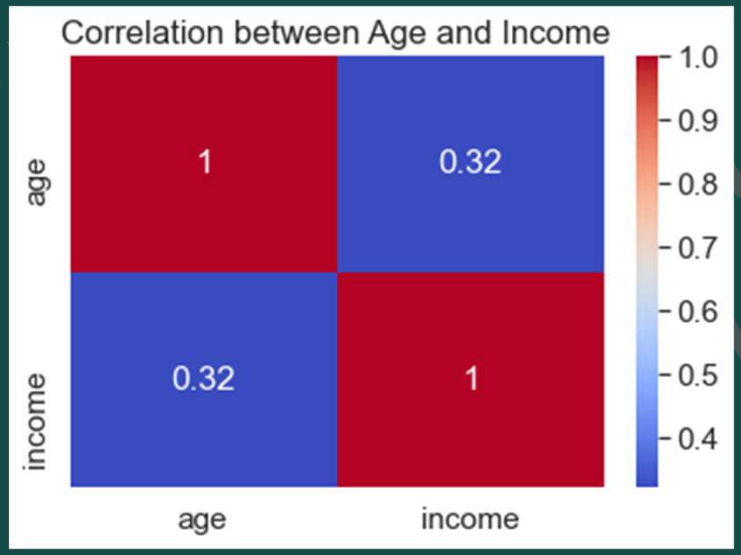
```
There are 17000 unique Starbucks members in this dataset.
```

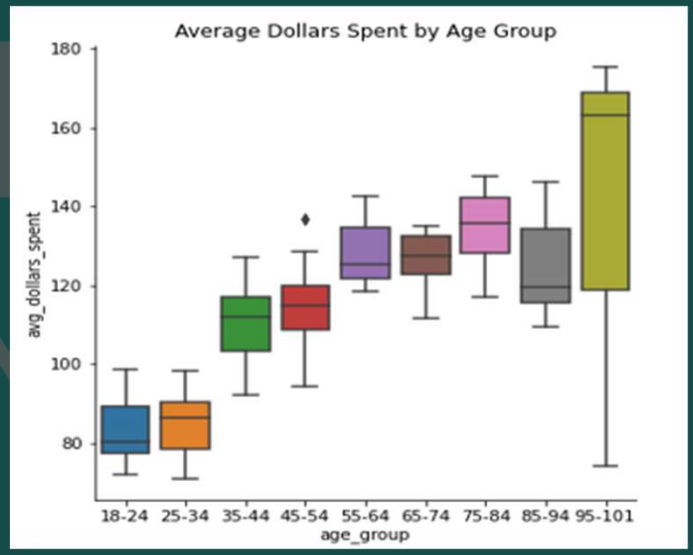
```
1 print("In aggregate, customers spent", round(transcript['amount'].sum(),2), "dollars in this one month!")
```

```
In aggregate, customers spent 1775451.97 dollars in this one month!
```

Understand the Customer:

How much have these members spent at Starbucks by age group?

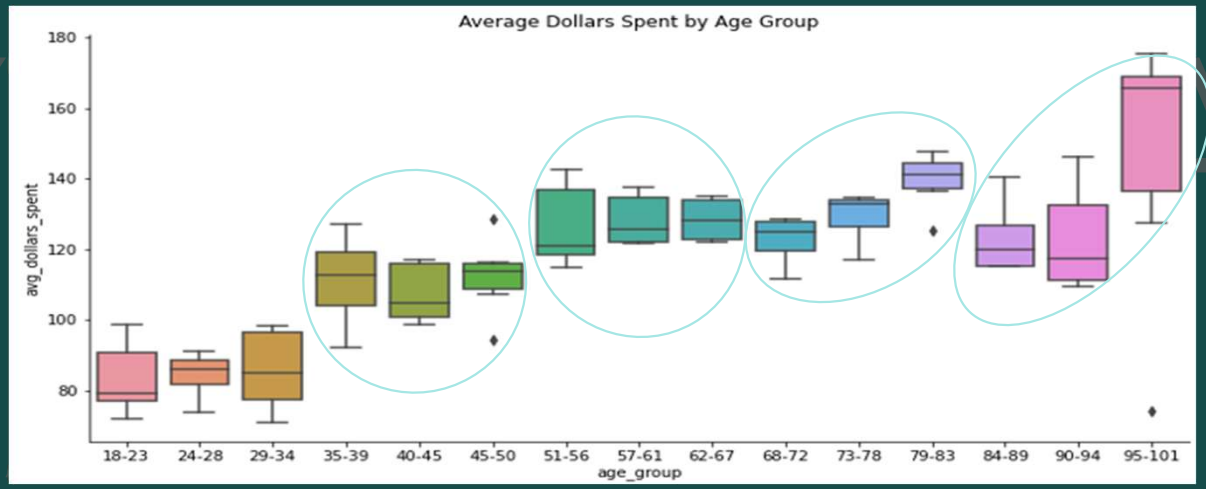




Customers spend more at Starbucks with age as income increases.

Understand the Customer:

How much have these members spent at Starbucks by age group?



Using **Sturges rule from Statistics** to determine the optimal number of bins ($\log_2 N + 1$), 15 bins were prescribed.

Recommendation: **Target age groups 35-50, 51-67, 68-83, 84-101** with reward offers that resonate with them in their stage of life to increase their propensity to spend because they command higher spending power.

Understand the Customer:

What is the spend range of Starbucks members in this dataset?

How many and what percentage of members have spent AND not spent this month?


```
print("Their spend ranges from", transcript.groupby('person')['amount'].sum().min(),\
"to",transcript.groupby('person')['amount'].sum().max(), "dollars in this one month.")
```

```
Their spend ranges from 0.0 to 1608.69 dollars in this one month.
```

```
total_members=len(pd.unique(transcript['person']))
paying_members=transcript.groupby('person')['amount'].count()\
               .sort_values(ascending=False).astype(bool).sum(axis=0)
non_paying_members=total_members-paying_members

print("There are",paying_members,"Starbucks members that have spent this month and \n",\
      non_paying_members,"members that have not spent this month.\n")
print("In other words, that is", round(paying_members/total_members*100,1), \
      "percent of the population that are active users \nand",\
     round(non_paying_members/total_members*100,1), "of the population that are inactive users")
```

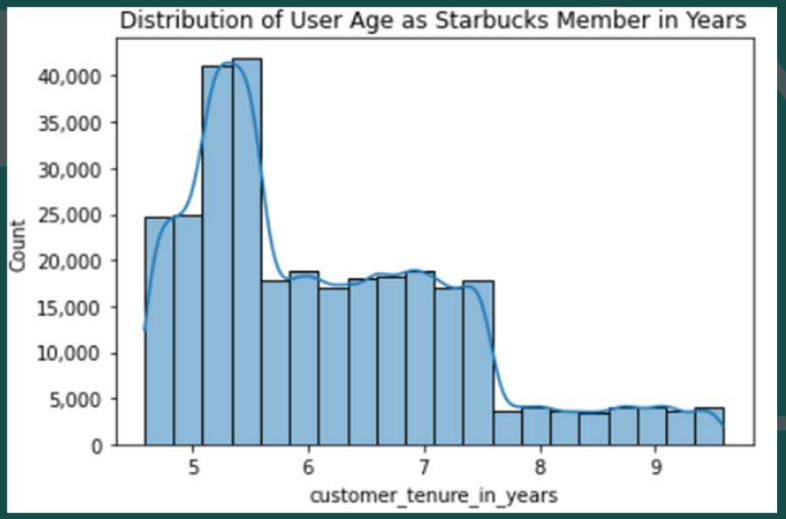
```
There are 16578 Starbucks members that have spent this month and
 422 members that have not spent this month.

In other words, that is 97.5 percent of the population that are active users
and 2.5 of the population that are inactive users
```

Opportunity to activate 2.5% of inactive members if that is a priority.
One tactic: Offer a free good or pair a free good with a conditional action.

Understand the Customer:

What is the distribution of User Tenure in this dataset (as of today)?

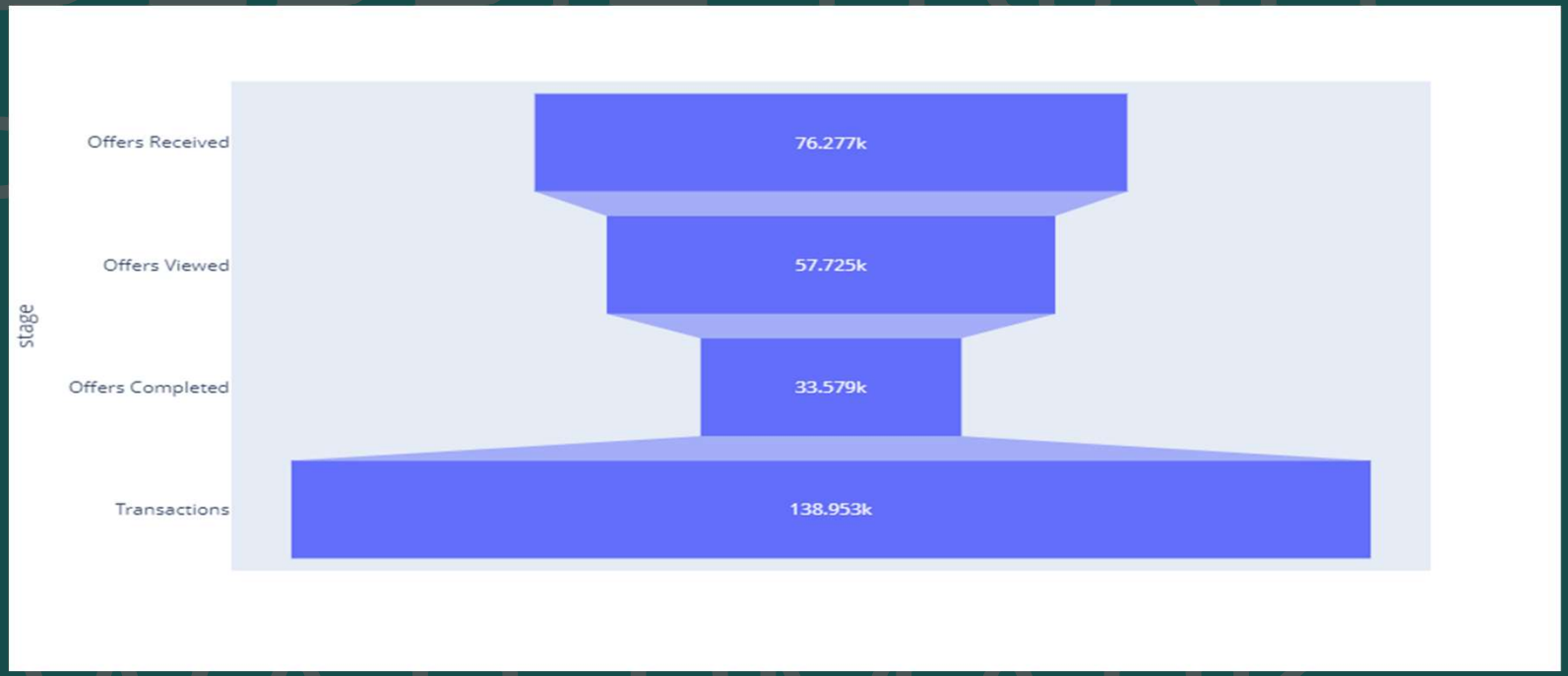



```
customer_tenure['customer_tenure_in_years'].round(1).describe()
```

```
count    306534.000000
mean          6.153731
std           1.155428
min           4.600000
25%           5.200000
50%           5.900000
75%           6.900000
max           9.600000
Name: customer_tenure_in_years, dtype: float64
```

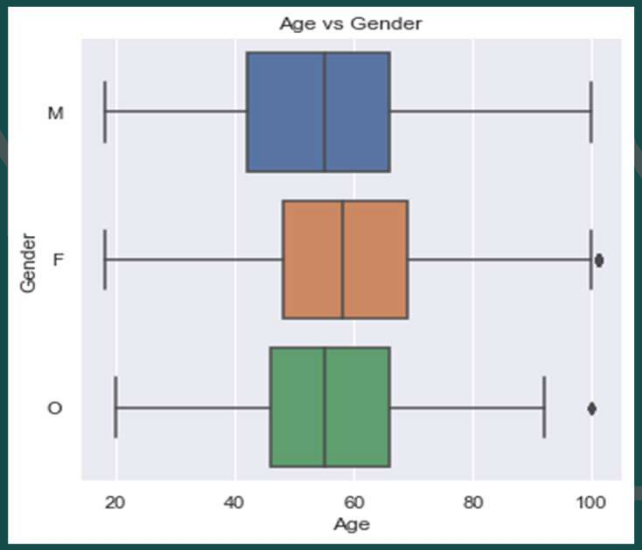
The average user tenure is 6 years, ranging from 4.5 and a half to 9.5 years. This may be an important feature for customer segmentation.

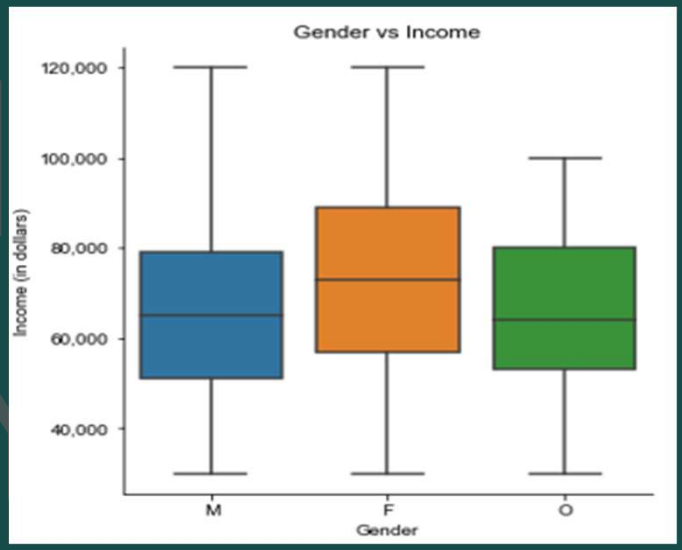
How many actions were taken in the marketing funnel in this dataset?



Number of transactions are higher than offers received.
While reward offers influences user behavior, some customers purchase independent of offers.

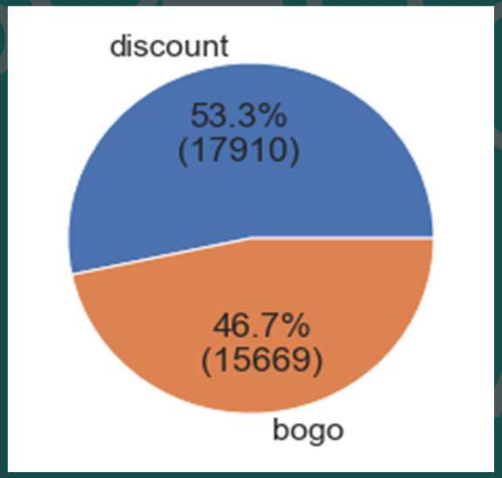
Which customer personas complete reward offers?





Starbucks reward offers can appeal to anyone across demographic categories.

Of rewards offers completed, what kind of offer is it?



Discount offers perform relatively better than BOGO rewards offers.

What may explain this is customers don't need two things when they go to Starbucks.

Their motivation may be to go to buy one unique item while also experiencing Starbucks as a 'Third Place'.

Which variant of the reward type has the most completed offers?

	offer_type	difficulty	reward	duration	ranked_num_offers_completed
0	discount	10.0	2.0	10.0	5317
1	discount	7.0	3.0	7.0	5156
2	bogo	5.0	5.0	7.0	4354
3	bogo	5.0	5.0	5.0	4296
4	discount	10.0	2.0	7.0	4017
5	bogo	10.0	10.0	7.0	3688
6	discount	20.0	5.0	10.0	3420
7	bogo	10.0	10.0	5.0	3331

offer_type	difficulty	reward_x	duration	num_offers_completed
discount	20.0	5.0	10.0	3420
bogo	10.0	10.0	7.0	3688
			5.0	3331
discount	10.0	2.0	10.0	5317
			7.0	4017
	7.0	3.0	7.0	5156
bogo	5.0	5.0	7.0	4354
			5.0	4296

Offers with the same difficulty and reward performed better with a longer offer duration.

This may be happening because the longer the offer is available, the likelihood the user has enough days to prioritize it while having urgency to act on it and complete the rewards offer.

Which demographic completes the most offers and what type of reward offer variant appeals to them most?

	gender	offer_type	difficulty	reward	duration	percent of offers completed/offers received
0	F	bogo	5.0	5.0	5.0	69.8
1	F	bogo	5.0	5.0	7.0	72.2
2	F	bogo	10.0	10.0	5.0	63.8
3	F	bogo	10.0	10.0	7.0	67.5
4	F	discount	7.0	3.0	7.0	81.2
5	F	discount	10.0	2.0	7.0	67.9
6	F	discount	10.0	2.0	10.0	82.2
7	F	discount	20.0	5.0	10.0	60.8
8	M	bogo	5.0	5.0	5.0	56.7
9	M	bogo	5.0	5.0	7.0	55.4
10	M	bogo	10.0	10.0	5.0	40.1
11	M	bogo	10.0	10.0	7.0	45.3
12	M	discount	7.0	3.0	7.0	68.0
13	M	discount	10.0	2.0	7.0	52.4
14	M	discount	10.0	2.0	10.0	70.3
15	M	discount	20.0	5.0	10.0	42.3

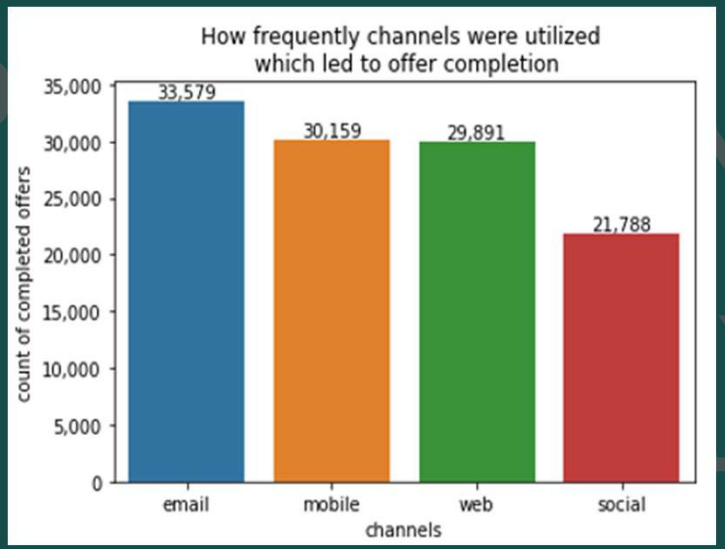
	gender	offer_type	difficulty	reward	duration	percent of offers completed/offers received
0	F	discount	10.0	2.0	10.0	82.2
1	F	discount	7.0	3.0	7.0	81.2
2	F	bogo	5.0	5.0	7.0	72.2
3	M	discount	10.0	2.0	10.0	70.3
4	F	bogo	5.0	5.0	5.0	69.8
5	M	discount	7.0	3.0	7.0	68.0
6	F	discount	10.0	2.0	7.0	67.9
7	F	bogo	10.0	10.0	7.0	67.5
8	F	bogo	10.0	10.0	5.0	63.8
9	F	discount	20.0	5.0	10.0	60.8
10	M	bogo	5.0	5.0	5.0	56.7
11	M	bogo	5.0	5.0	7.0	55.4
12	M	discount	10.0	2.0	7.0	52.4
13	M	bogo	10.0	10.0	7.0	45.3
14	M	discount	20.0	5.0	10.0	42.3
15	M	bogo	10.0	10.0	5.0	40.1

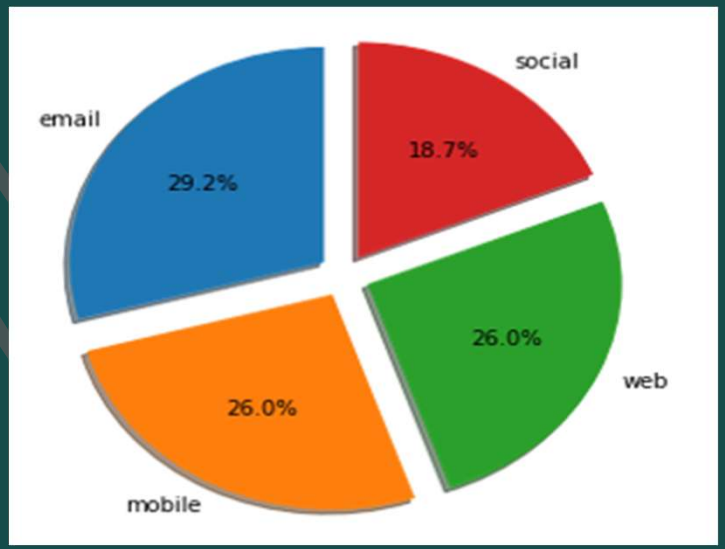
The top 10 performing rewards offers are typically completed by females, have a 7-10 duration to complete, and are discount offers.

The best performing BOGO offer by gender is valued at $5.

The bottom performing offers by gender are $20 difficulty discount & $10 BOGO offers.

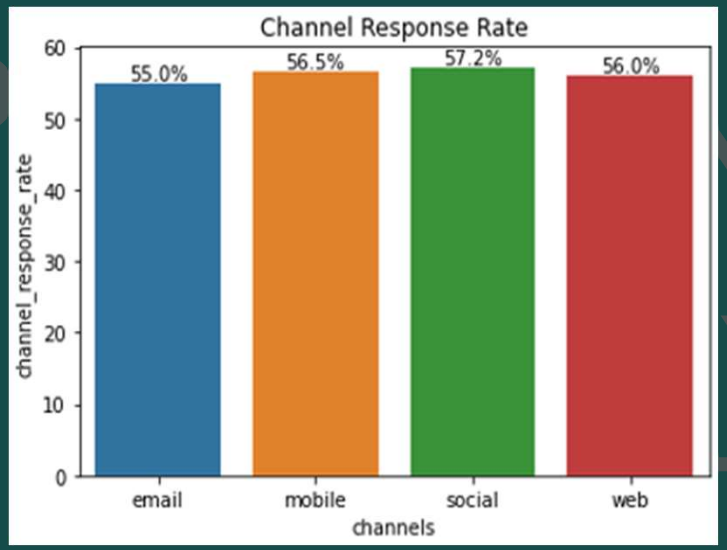
Which channels were used most frequently which led to offer completion?

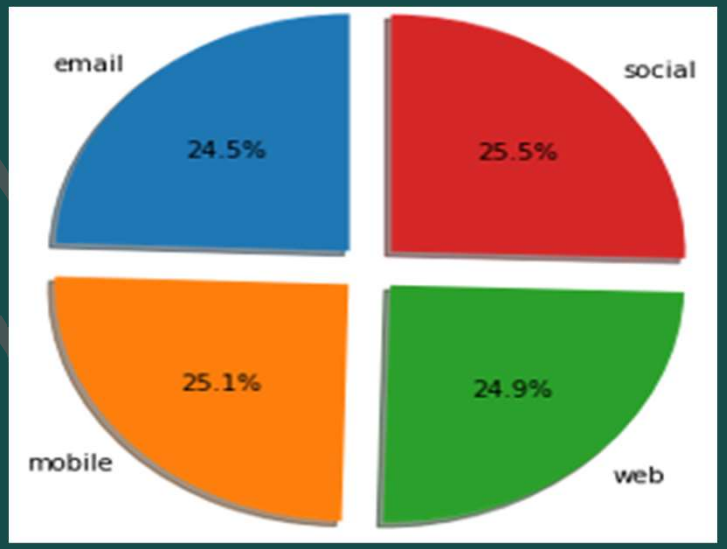




Emails were used most frequently to complete offers, followed by mobile & web, and then social.

Which channel has been most effective in driving BOGO/Discount reward offers completion?





It's essentially a tie.

Each channel has similar response rates from the user of 55-57%.
All channel strategies appear to be effective.

Direct Attribution:
What is the rewards offer response rate?

offer_type	offers_received_response_rate	offers_viewed_response_rate	pct_diff
bogo	51.4	61.6	10.2
discount	58.6	83.5	24.9

Response rate is a behavioral metric to measure success of Starbucks reward offers strategy:

A. Direct Response Rate = Number of Offers Completed / Number of Offers Received

B. Direct Response Rate = Number of Offers Completed / Number of Offers Viewed

Direct Attribution:
What is the rewards offer response rate?

offer_type	offers_received_response_rate	offers_viewed_response_rate	pct_diff
bogo	51.4	61.6	10.2
discount	58.6	83.5	24.9

If you can get customers to view the offer,

the response rate improves by 25% for discount offers & 10% for BOGO offers.

Bottom Line: Get users to be positively engaged in the long run!

Recap of Insights

The trend is **customers spend more at Starbucks with age** as income increases.

2.5% of members are inactive. A tactic to activate them is by offering a complimentary item that can be contingent on a user purchase action.

Offers with the same difficulty & reward performed better with a longer offer duration.

While email is the most frequently utilized channel, **all channels** (email, mobile, web, social) **exhibit similar response rates from the user of 55-57%.** Continue with current channel strategy as it is.

Discount offers perform relatively better than BOGO rewards offers. If you can **motivate users to view the offer,** the response rate improves by 25% for discount & 10% for BOGO offers.

Marketing Strategy Actionable Recommendations:

Personas to Target & Strategy for Offer Type Variants

For future reward promotional strategy, **target age groups 35-50, 51-67, 68-83, 84-101** because they tend to spend more at Starbucks than younger age groups. One suggestion is to tailor rewards that resonate with them in their stage of life to increase their propensity to spend.

Continue to target female consumers by promoting Starbucks Discount reward offers, with a **7-to-10-day duration** to complete them.

Continue to promote the best performing BOGO offer by gender, which has a $5 difficulty and $5 reward.

Consider reducing or discontinue $20 difficulty discount offers and $10 BOGO offers, which does not as well relative to existing reward offers because of its difficulty to complete. Note: More analysis can be done to determine total transactions to credit each reward offer type variant.

What's next?

Roadmap ahead

Next Steps to Improve Marketing Strategy via ML

1. Do customer segmentation and clustering based on key features (Note: Customer segmentation can help suggest similar reward offers to users that haven't seen that reward offer based on their cluster).
2. Create attribution rules for direct and indirect attribution to display as key metrics.
3. Create an R Shiny dashboard to display key metrics.
4. Feature engineer to enhance dashboard metrics, such as user tenure and average time to complete an offer.
5. Productionize model.


The way to get started is to quit talking and begin doing.
Walt Disney

Thank you

Debbie Trinh

https://www.linkedin.com/in/debbietrinh/

New York City Data Science Academy

Jan 2023 Cohort