Starbucks Capstone Project Report

Overview

This data set contains simulated data that mimics customer behaviour on the Starbucks rewards mobile app. Once every few days, Starbucks sends out an offer to users of the mobile app. An offer can be merely an advertisement for a drink or an actual offer such as a discount or BOGO (buy one get one free). Some users might not receive any offer during certain weeks.

Not all users receive the same offer, and that is the challenge to solve with this data set.

The task is to combine transaction, demographic and offer data to determine which demographic groups respond best to which offer type. This data set is a simplified version of the real Starbucks app because the underlying simulator only has one product whereas Starbucks sells dozens of products.

Every offer has a validity period before the offer expires. As an example, a BOGO offer might be valid for only 5 days. You'll see in the data set that informational offers have a validity period even though these ads are merely providing information about a product; for example, if an informational offer has 7 days of validity, you can assume the customer is feeling the influence of the offer for 7 days after receiving the advertisement.

Dataset

The data is contained in three files:

- portfolio.json - containing offer ids and meta data about each offer (duration, type, etc.)
- profile.json - demographic data for each customer
- transcript.json - records for transactions, offers received, offers viewed, and offers completed

Here is the schema and explanation of each variable in the files:

portfolio.json

- id (string) - offer id
- offer_type (string) - type of offer ie 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)

profile.json

- 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

transcript.json

- event (str) - record description (ie transaction, offer received, offer viewed, etc.)
- person (str) - customer id
- time (int) - time in hours since start of test. The data begins at time t=0
- value - (dict of strings) - either an offer id or transaction amount depending on the record

Business Problem

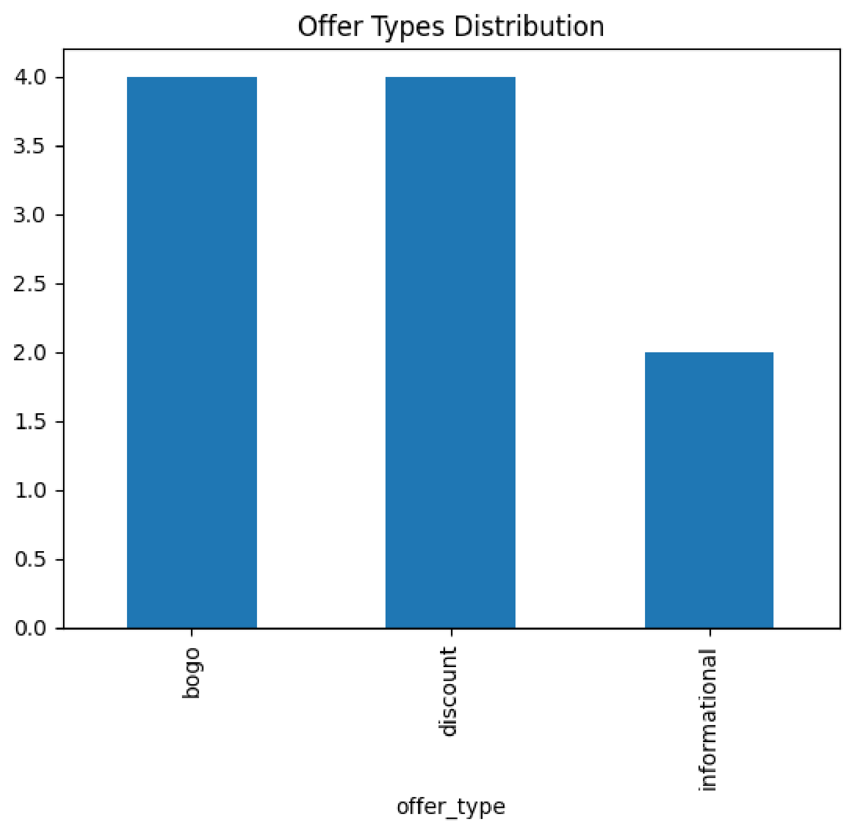
Starbucks want to personalise their marketing more by understanding which types of customers respond best to which types of offers. So, what we need to understand is, which off could be sent to a particular person which is likely to increase engagement and activity.

The key challenge is that not all users react in the same way. For example, some customers may ignore offers, while others will be influenced by them. Some users may complete an offer but might have never viewed it – some just completed it because they wanted to make an order and didn't notice the offer. For this reason and others, the data requires cleaning. My aim is to understand how each demographic group across, gender, age group and income bracket react towards different offer type. I want to be able to analyse to understand each group, build a model and share suggestions on what Starbucks could do moving forward.

Data set Analysis

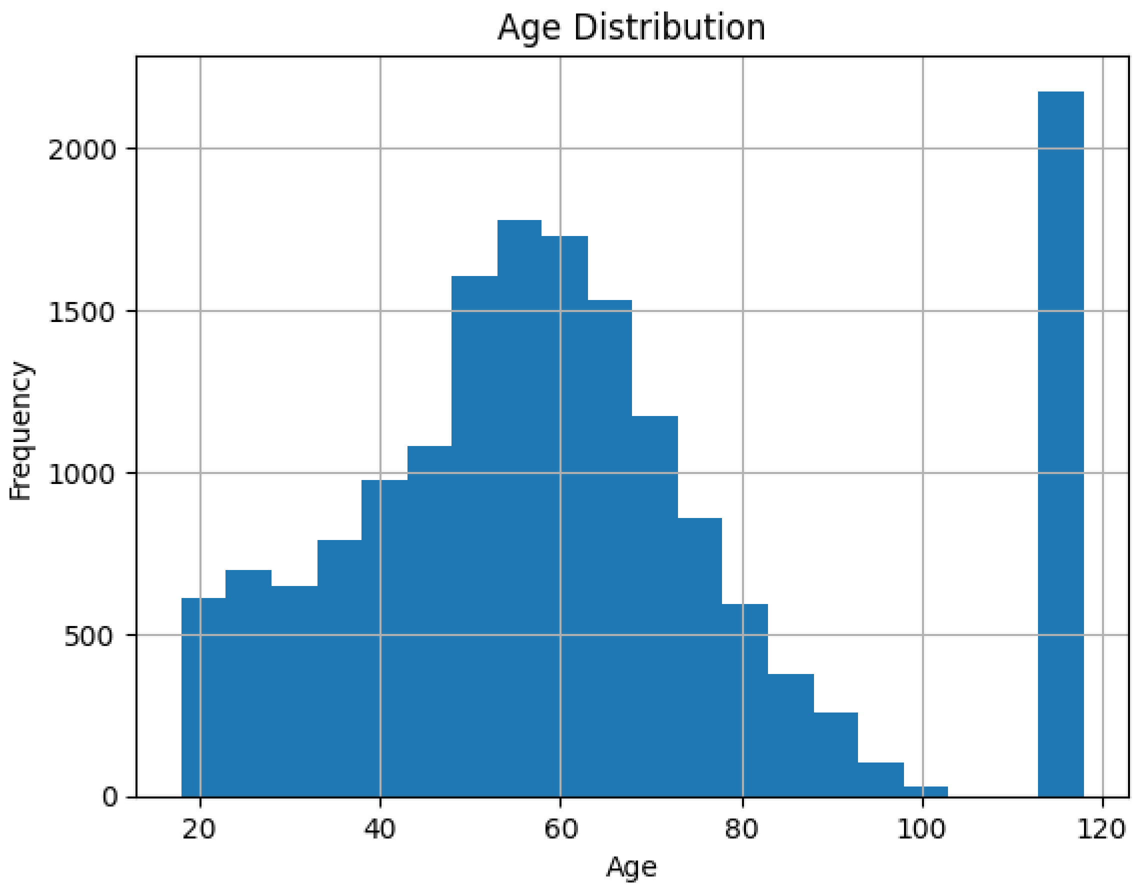
Portfolio

The data set consists of no null values which will help massively with the cleaning and merging process. Offers distributed are higher for BOGO and Discount.



Profile

Upon analysis there appears to be an anomaly with customers listed as age 118. This value occurs with unusually high frequency then expected, particularly when compared to 101. Additionally, the count for age 118 matches with the count of missing values for gender and income. As such, these records are likely invalid and may be best to remove from the dataset.



When looking at the missing values, all the missing values come from age 118 for gender and income. This further highlight that it is best to remove this age from the dataset to continue with analysis and recommendations.

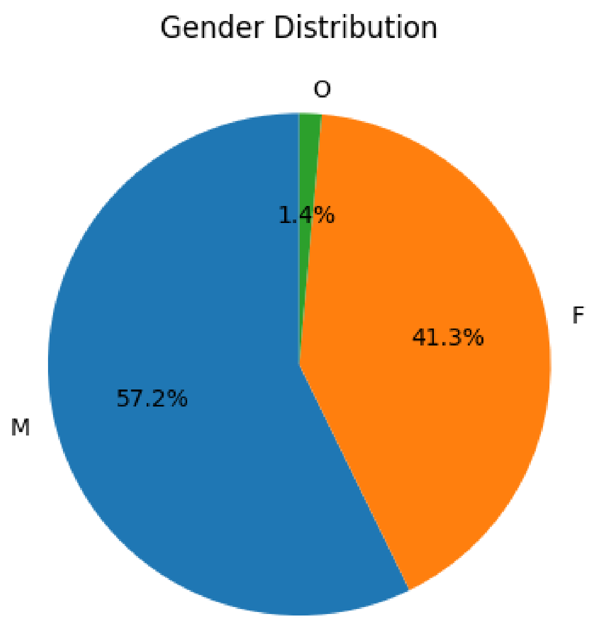
```python
# Filter rows where age is 118
age_118_rows = profile[profile['age'] == 118]

# Check missing values in gender and income for these rows
print(age_118_rows[['gender', 'income']].isnull().sum())
```

✓ 0.0s Python

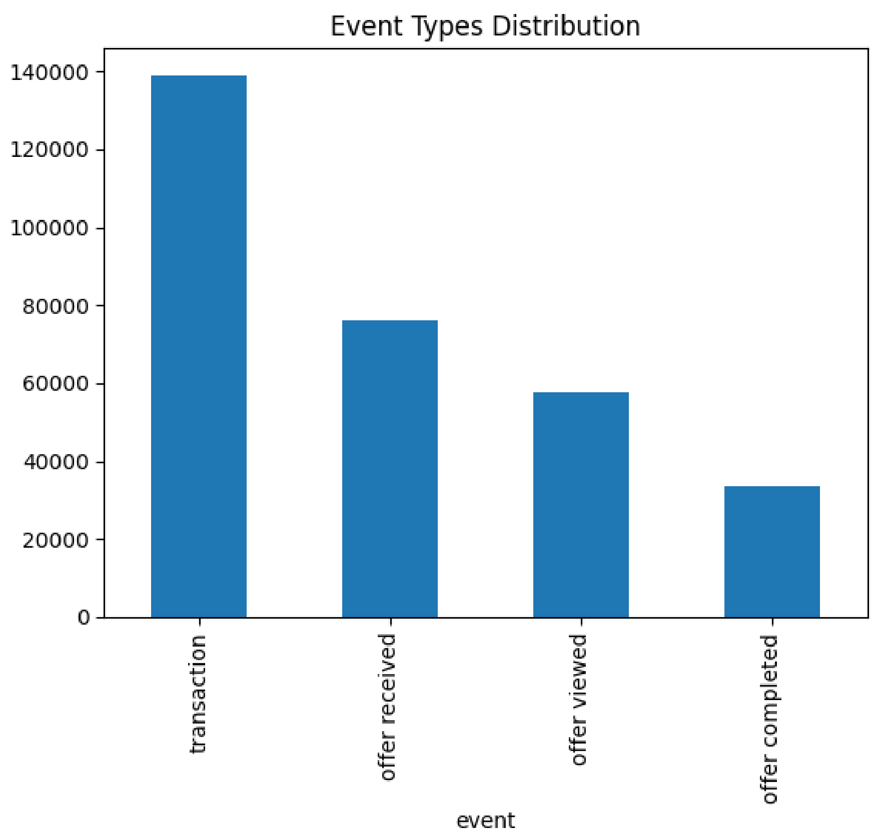
```
gender    2175
income    2175
dtype: int64
```

We have more males then female in the dataset.



Transcript

The data has no null values. We can see that there are more transactions, and it reduces for each event, with offer completed being the lowest.



Cleaning and merging the data set

We performed cleaning and data preparation to get to a place where we have a dataset that we can work with for our analysis and to make recommendations. The main steps taken are outlined below.

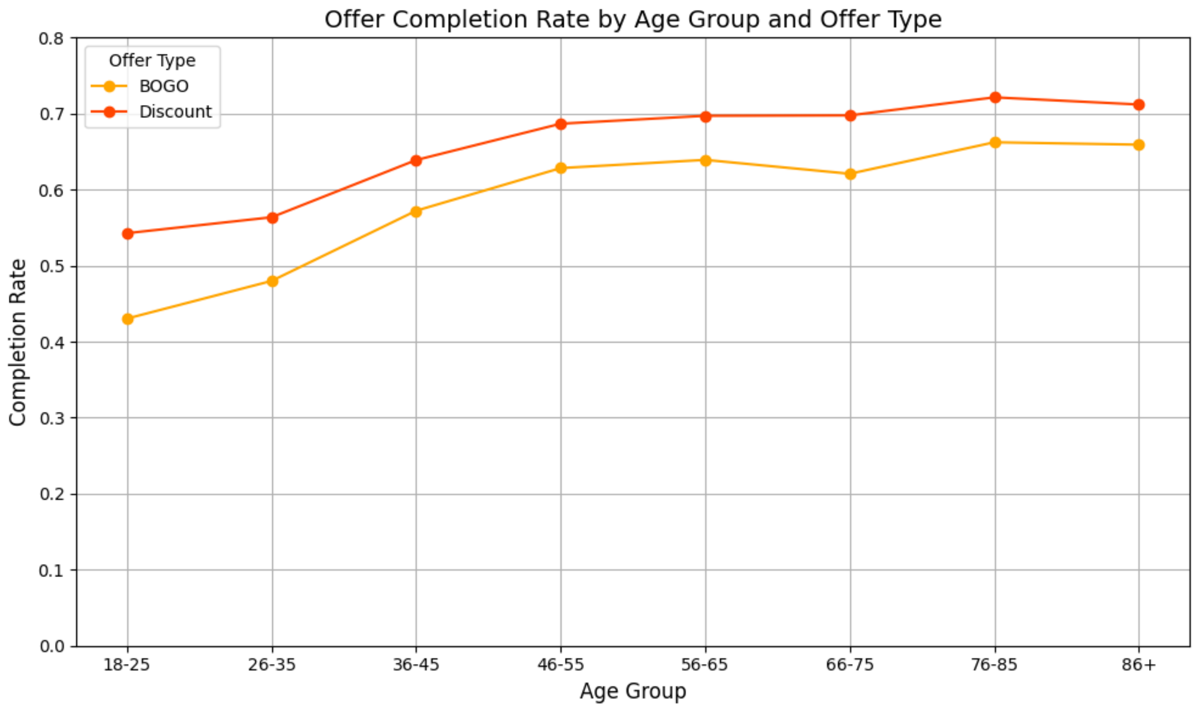
- Demographic data cleaning – We removed the 2175 customer profiles that had missing gender and income. The age looked like an anomaly and the age distribution look more realistic once removed. Once the data set was cleansed it consisted of 14,825 customers. Once cleaned, there were no nulls found and all customers has a valid age, gender and income. We also converted become_member_on filed to a date field to show each customers membership tenure in years.
- Offer portfolio processing – we rename the columns which was to help with clarity and merging. For example, id to offer_id. We expanded the channels into separate binary columns for each channel. We one hot encoded the offer_type to help with analysis. After the clean, no values were missing from the dataset.
- Transcript – we split the value column into offer_id and amount which would help with filtering after merge. We cleaned names and for each event type ("offer received", "offer viewed", "offer completed", "transaction") we created a flag. We removed event related to the dropped age column, so that the transcript column now has only 14,825 valid customers.
- Merged dataset – we merged the data set. This created a data frame that was able to show each offer sent to customers, with the customers demographic and whether they have viewed and/or completed the offer. To do this, we grouped transcript events by offer and person. For each offer received, we looked ahead in the transcript for a corresponding viewed or completed event by the same customer for the same offer within the offer's validity period. We then marked the offer as viewed/completed as appropriate. If an offer was completed without a view, it's marked completed=1, viewed=0. If it was viewed and not completed by expiry, that's viewed=1, completed=0. Transactions that were unrelated to any offer were ignored in this merge (though they could be analysed for insight on baseline spending).

After these steps, we had a clean, consolidated dataset of offer outcomes (one row per offer per customer) ready for analysis. This structured format was crucial for calculating response rates and training the predictive model.

Key findings from analysing the datasets

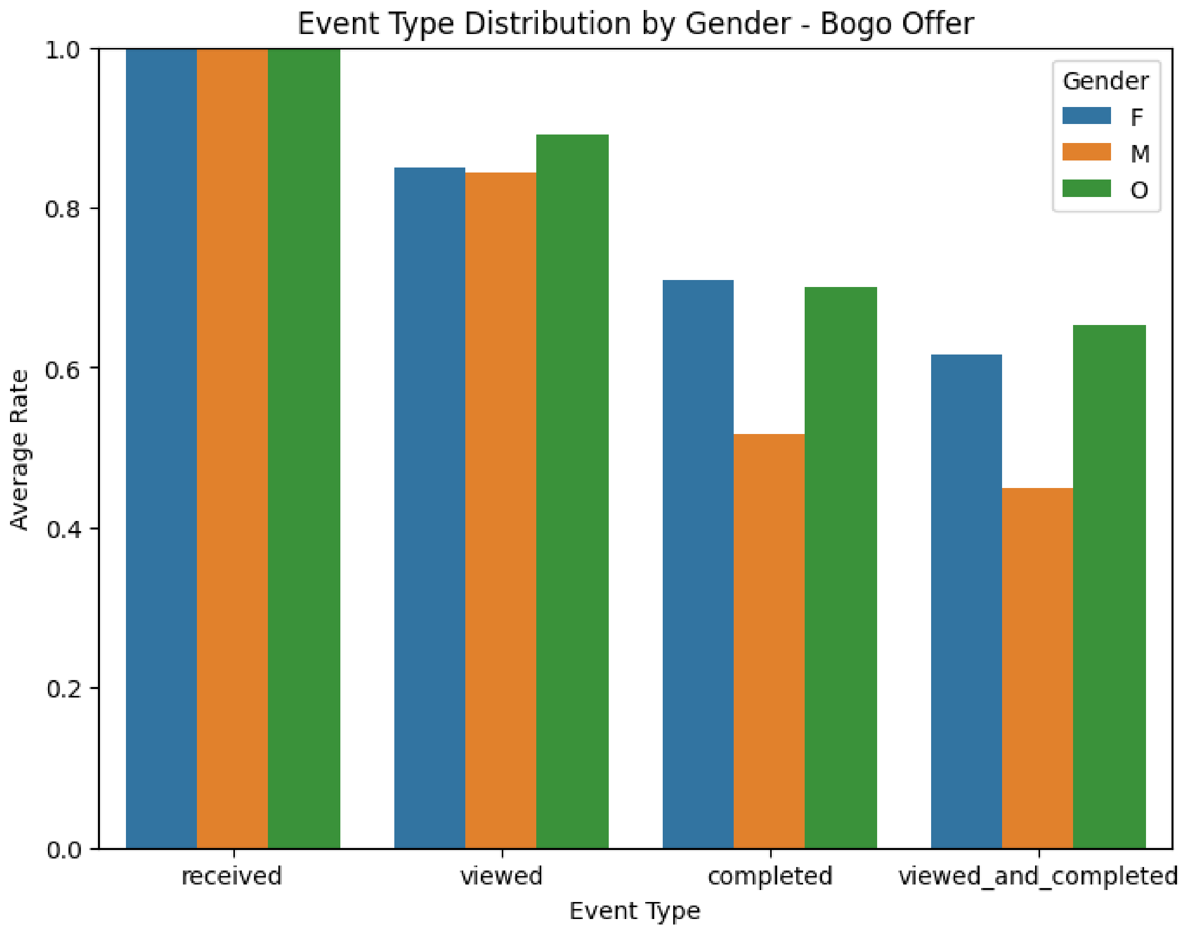
By analysing the dataset, we were able to find several insights about customer segments and behaviour.

- **Age Groups:** Age is a strong predictor of offer response. **Older customers respond at much higher rates to offers than younger customers.** We saw a clear upward trend: customers under 30 had the lowest completion rates for offers, while customers in their 50s, 60s, and beyond had the highest completion rates. The below chart shows this trend for two offer types:

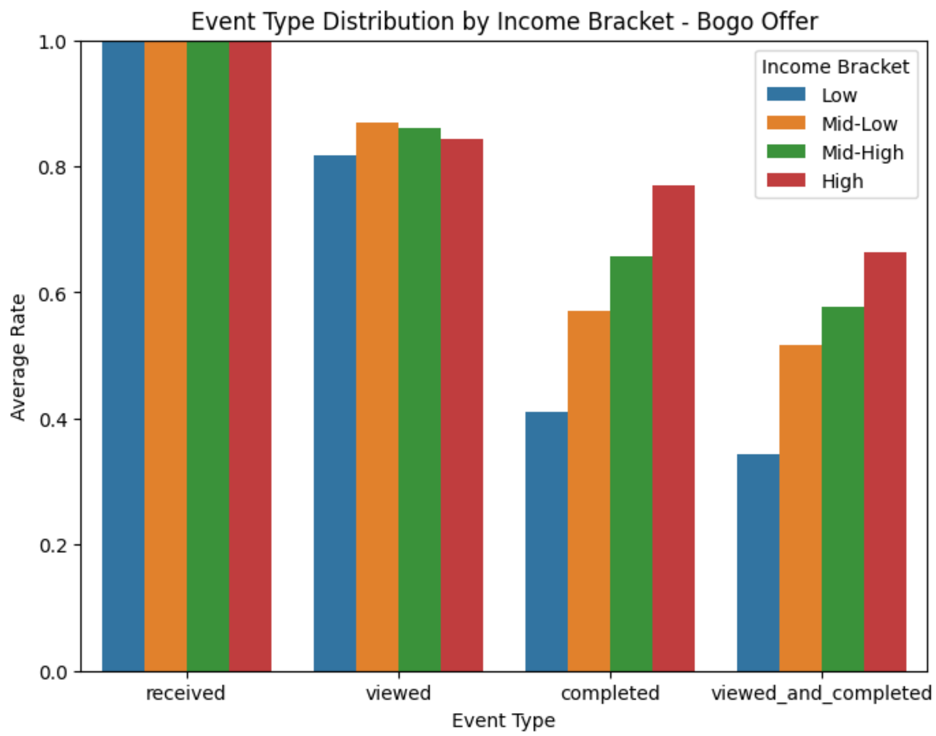


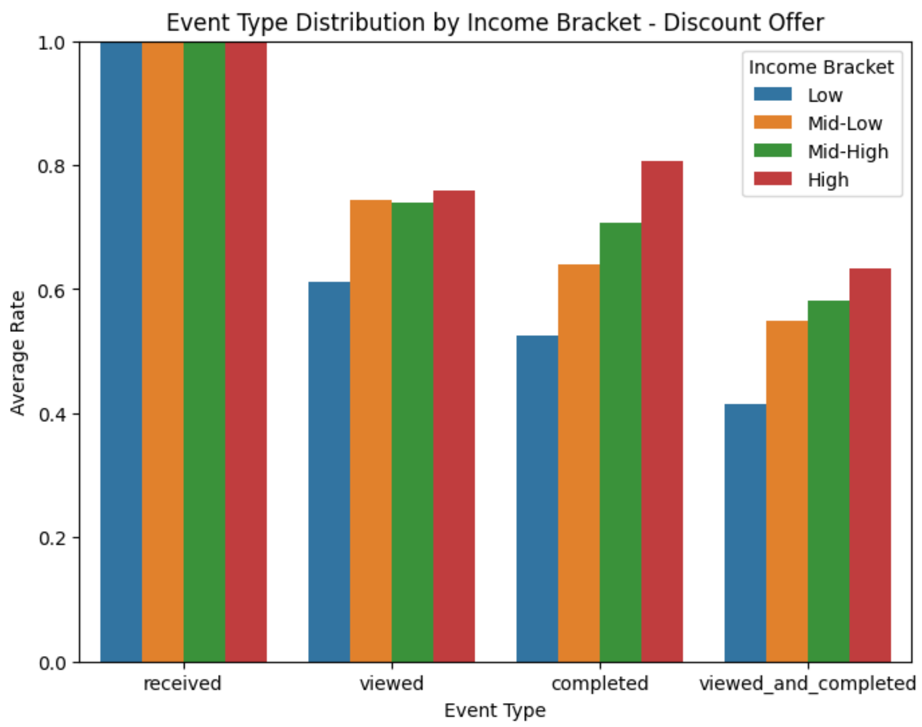
This means promotional offers resonate much more with middle-aged and senior customers. A 55-year-old customer is far more likely to use that "Buy One Get One" coupon than a 25-year-old customer. Younger customers might require different engagement strategies - they might be less interested in traditional discounts or possibly not as financially able to take advantage of them (we did see the average income increases as age increases).

Gender: Female customers were more receptive to offers than male customers. When segmenting offer outcomes by gender, females were more likely to complete the offer then men. Men comprised a majority of the customer base in this dataset but were still less likely to take up the offers. This is clearer when the looking at BOGO offers (illustrated in the chart below). Both groups have viewed the offers, but Females have viewed and completed a lot more than men. As Females are more likely to use the offer, it may be an option to increase female targeted promotions or additional way to encourage males.



Income Levels: Higher income customers tend to complete offers at greater rates. We observed that customers in the top income bracket had a higher proportion of completed offers compared to those in the lowest bracket. This trend is intuitive: higher income implies more discretionary spending, making it easier to meet offer requirements (like spending thresholds for a discount). Lower-income customers may skip some offers if they require spending beyond their budget. For Starbucks, this could mean that premium offers or upsell campaigns work well in wealthier demographics, whereas value-oriented offers (lower minimum spend or longer duration) might be needed for lower-income segments to participate.





- **Offer Type Performance:** Comparing offer types:
  - **Discount offers** had the highest overall completion rate. They seem broadly effective across demographics.
  - **BOGO (Buy-One-Get-One) offers** had a slightly lower completion rate. They are still effective but appear slightly less so than discounts. Certain groups (e.g., younger customers) particularly showed low uptake on BOGOs, possibly because they might not want two of the same items or don't have someone to

share with. However, older customers responded nearly as well to BOGOs as to discounts, narrowing the gap in those segments.

  - **Informational offers** had no direct reward of completion metric although they were still viewed. While they cannot be completed like discount or BOGO offers, the view rate suggests they may still capture customer attention.

- **Behavioural Insight – Viewing vs. Completing:** A critical behavioural finding was the importance of the customer viewing an offer. A customer is more likely to complete an offer when viewed, although there were cases where customers were completing offers in which they had not viewed. Effective marketing would be ensuring the offers are viewed to influence customer behaviours. So, an action on thus would be to increase offer view rates. This could be done by sending more reminders, making it easier and simpler to see or highlighting offers in the app so it is hard to miss.

These findings paint a clear picture: the **ideal target for Starbucks offers is an older, higher-income, female customer who actively uses the app**. That doesn't mean other groups should be neglected, but Starbucks might tailor strategies to improve uptake in under-performing segments (younger, lower-income, male customers).

Model Training and Evaluation Summary

To complement the exploratory analysis, we developed a machine learning model to predict whether a customer would **view and complete an offer**. This model helps Starbucks better understand which customer-offer combinations are likely to succeed, enabling more targeted and cost-effective marketing.

1. Features Used

The model was trained using a combination of offer characteristics and customer demographics:

- **Offer Type**: Categorical (BOGO, Discount, Informational)
- **Viewed**: Binary flag indicating whether the customer viewed the offer
- **Age Group**: Categorized age bins
- **Gender**
- **Income Bracket**: Derived from income distribution

All categorical variables were one-hot encoded to prepare for model training.

2. Model Choice

We used a **Random Forest Classifier** due to its ability to:

- Handle both categorical and numerical data
- Capture nonlinear relationships
- Provide feature importance insights
- Remain interpretable for business use

To address class imbalance, we enabled class_weight='balanced'.

3. Training Procedure

The dataset was split into 70% training set and 30% test set. Stratification was used to preserve the target class distribution. Default Random Forest parameters were used to establish a baseline model. No hyperparameter tuning was performed at this stage.

The model achieved strong results

```
=== Classification Report ===
```

	precision	recall	f1-score	support
0	0.95	0.73	0.83	9517.00
1	0.72	0.95	0.82	7050.00
accuracy	0.82	0.82	0.82	0.82
macro avg	0.84	0.84	0.82	16567.00
weighted avg	0.85	0.82	0.82	16567.00

```
=== Confusion Matrix ===
```

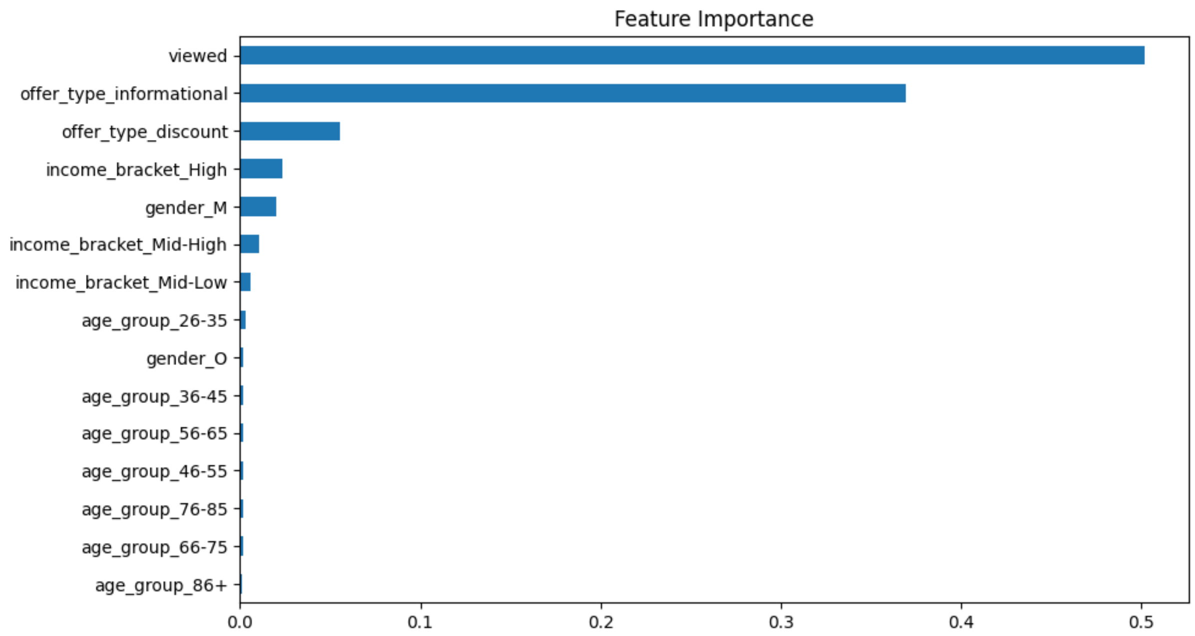
	Predicted: No	Predicted: Yes
Actual No Offer	6944	2573
Actual Completed	362	6688

The model is very effective at identifying customers who will respond to offers (**high recall for completions**).

Some false positives exist (predicting a completion that doesn't happen), but this is acceptable from a business perspective since the cost of sending an unused offer is low.

5. Feature Importance

The most influential feature in predicting success is viewed.



Model Development and Iteration Process

To find the best-performing model, we started with a baseline approach and gradually improved it through experimentation. We trained a model using default hyperparameters. While strong overall, we observed some over-prediction of completions (leading to false positives). We accepted this trade-off due to the low cost of sending unused offers.

We considered additional features such as time since membership, offer duration and reward, but decided to focus on offer type, gender, age group and income bracket. This was to avoid multicollinearity/redundant features. Including them could confuse the model with overlapping info. This makes the model easy to explain and still answers the question required.

Random Forest vs Logistic Regression

We selected the Random Forest Classifier as our primary model for predicting whether a customer would view and complete a marketing offer. Although Logistic Regression yielded similar performance (approximately 82% accuracy), Random Forest was chosen due to its greater flexibility and ability to handle complex patterns in the data. Unlike Logistic Regression, which assumes a linear relationship between features and the target variable, Random Forest does not make those assumptions. This is particularly valuable in modelling customer behaviour, where interactions between demographic variables (like age, income, and gender) and offer types (such as BOGO or discount) may influence outcomes in nonlinear ways.

Additionally, Random Forest provides clear feature importance scores, helping us identify the most influential factors in offer completion. In this case, variables such as viewed, offer type, and income bracket emerged as top predictors, aligning well with insights from our exploratory analysis. While Logistic Regression offers simplicity and interpretability through coefficients, Random Forest balances predictive power with interpretability. For these reasons, Random Forest was selected as the final model, offering a strong blend of performance, robustness, and actionable insights.

Evaluation Metrics & Justification

To assess the performance of our classification model, we used several metrics that reflect both the predictive accuracy and the practical impact of errors. Accuracy provides a general measure of how often the model correctly predicts offer completions and non-completions. However, because it treats all errors equally, accuracy can be misleading when classes are imbalanced. To gain deeper insight, we evaluated precision, recall, and the F1 score.

Precision is particularly important for understanding the cost of false positives — cases where the model predicts that a customer will complete an offer when they actually won't. While Starbucks can tolerate some false positives due to the relatively low cost of sending an unused offer, minimising them improves marketing efficiency. Recall, on the other hand, is even more critical. It measures the proportion of actual completed offers that the model successfully identifies. A high recall ensures that we are capturing as many potential responders as possible, which is vital for maximising the reach and impact of promotional campaigns.

The F1 score balances precision and recall into a single metric, making it particularly useful when both types of errors — false positives and false negatives — carry business consequences. In this case, a model that simply predicts "no completion" for everyone could still achieve high accuracy, but it would fail to identify customers who would actually respond. For this reason, we placed greater emphasis on recall and F1 score, especially for the positive class (completed offers), to ensure the model effectively identifies likely responders — even if it occasionally sends an offer to someone who does not act on it.

Business Recommendations

Based on the analysis, we propose the following actionable recommendations for Starbucks' marketing strategy:

1. **Target Demographics with Tailored Offer Types:**
    - **Older Customers:** They have the best completion rates. Continue sending them **discount offers**, as they particularly respond well to discounts. BOGO also works for them, so a mix is fine. Ensure this age group consistently receives promotional offers, as they are likely to redeem them and drive sales.
    - **Younger Customers:** Their low engagement suggests a need for rethinking the approach. Standard discounts/BOGOs aren't as effective here. So alternative strategies may be bast such as
        - Using more experiential **rewards** or **gamified** rather than straight money-off – something that creates excitement or social sharing might engage them more.
        - Use **social media integration** or app features to highlight offers.
        - If offering discounts, possibly lower the spending threshold for this group to make it more attainable, or pair offers with popular items among younger consumers.
2. **Gender-Specific Campaigns:**
    - **Female-Focused Offers:** Since women are more responsive, consider campaigns that specifically target female customers. For example, an offer timed around events that have higher female costumers.

- **Re-engage Male Customers:** For male customers who are less likely to use offers, try to identify what might motivate them:
    - It could be that the current offers (coffee, tea, certain food items) aren't as appealing. Perhaps introduce or highlight products that data shows men favour, and tie offers to those.
    - Another idea is to use **informational offers for men** that later lead into a targeted discount. For example, an informational offer about a new product, followed a week later by a personalised discount on that product might gradually hook those who didn't respond to generic offers.

3. **Improve Offer Visibility and Engagement:**
   Regardless of targeting, once an offer is sent, ensuring the customer notices it is crucial. Our analysis of view vs. completion underscored that an unseen offer is often a wasted offer. Starbucks should:
    - Send a **reminder notification** if an offer is nearing expiration and hasn't been used (and if the customer hasn't viewed it yet). A gentle nudge like "Don't miss out on your reward!" could improve view rates and hence completion.
    - Highlight offers in the mobile app interface more prominently. Perhaps a banner or a badge on the app icon indicating an unused offer. The more the offer stands out, the higher the chance the customer will open it.

4. **Further Segment and Experiment:**
   Starbucks can drill down further by conducting **A/B tests** within segments to see what specific tweaks improve response. For instance, within the under-35 group, test different offer messaging or different incentive levels to see what impacts them more, since we know this group is currently under-performing.

Conclusion

This analysis uncovered clear patterns in how different customer demographics respond to Starbucks' promotional offers. **Older, higher-income female customers** show the highest engagement, particularly with **discount offers**, while **younger, lower-income male customers** are the least responsive, especially to BOGOs.

To support these findings, we built a predictive model that accurately forecasts offer success using demographic and offer features. This model enables Starbucks to personalize marketing by matching the right offer to the right customer.

In short, middle-aged and older women with higher incomes are the most likely to complete offers—especially discounts—making them a high-value target group. Conversely, younger men are less likely to engage.

By leveraging both the insights and the model, Starbucks can **optimize promotions for maximum return**, focusing on proven responders while testing new strategies to convert under-engaged segments.