Title: Starbucks Capstone project:

Introduction Starbucks uses its rewards program to engage customers by sending tailored offers through its mobile app. These offers include buy-one-get-one (BOGO), discount, and informational promotions. The challenge lies in predicting which customers are most likely to respond to specific types of offers. In this project, we analyze customer demographics and transaction data to build a machine learning model that can improve offer targeting and boost engagement.

Data Overview The dataset is comprised of three main files:

1. **Portfolio**: This file contains details about the offers, including the type (BOGO, discount, or informational), the duration of the offer, and how difficult it is to complete.
2. **Profile**: Demographic data about the customers, including their age, gender, income, and membership status.
3. **Transcript**: A record of customer interactions with the app, including offers received, viewed, and completed, as well as transaction logs.

   By merging these datasets, we gain insights into customer behavior and preferences, which are essential for building a predictive model.

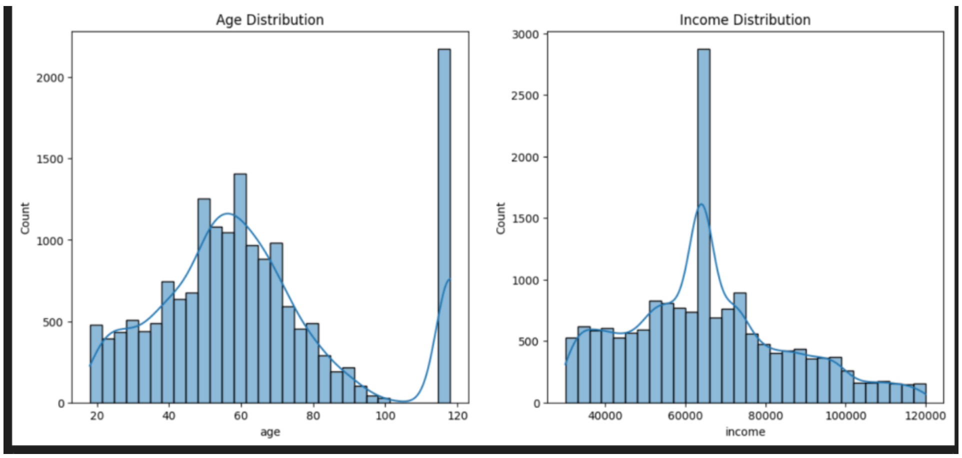
Methodology We approached the problem using a structured workflow:

1. **Data Preprocessing**:
   - Merging the datasets, cleaning null values, and converting timestamps into a usable format.
   - Encoding categorical variables (e.g., offer type and customer demographics) to make them suitable for machine learning models.
2. **Exploratory Data Analysis (EDA)**:
   - Analyzing trends and relationships between customer demographics and their response to offers.

- Visualizing offer completion rates by age, income, and gender.

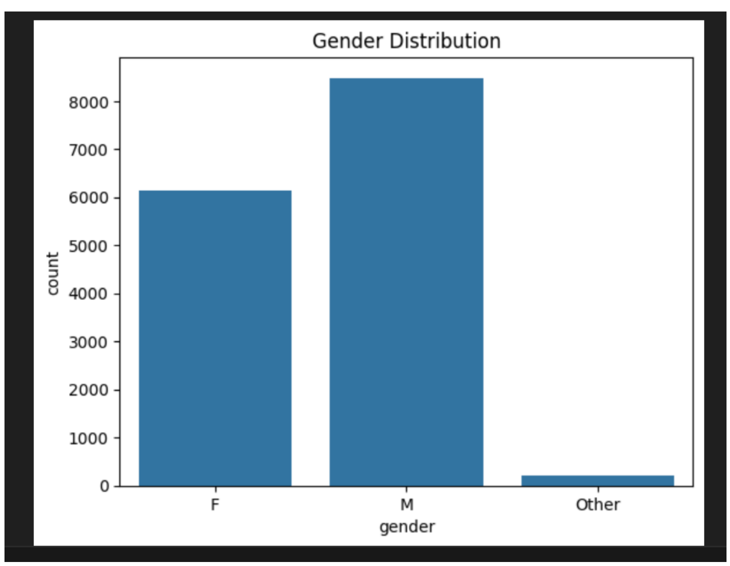
The distribution of age and income is visualized to understand the relationship between customer age and their income levels

This analysis is crucial for understanding which age groups are more likely to have higher income, which can influence how Starbucks targets different offers. Higher-income groups might respond better to premium offers, while younger, lower-income groups might be more responsive to discounts or promotional deals.



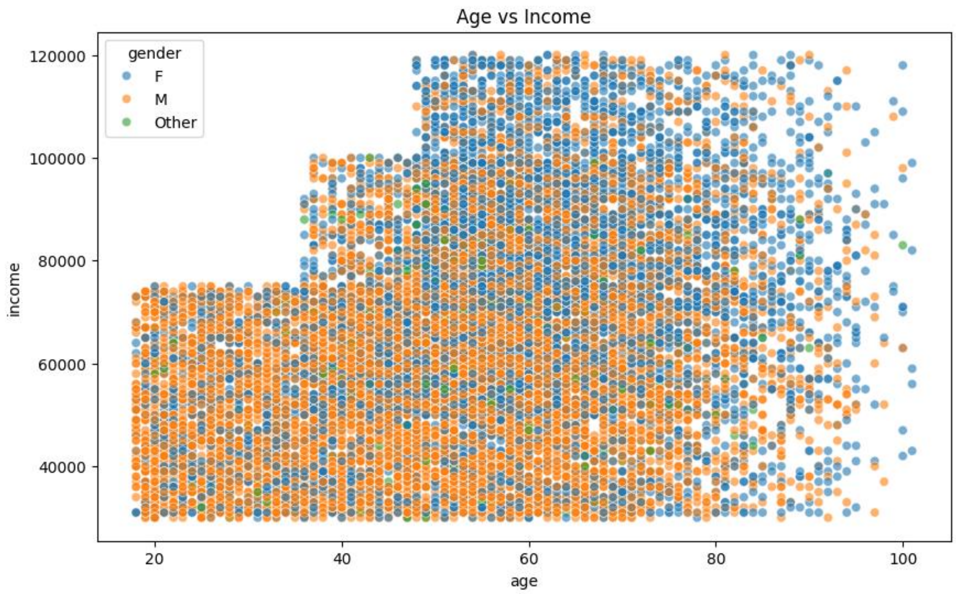
This bar plot illustrates the distribution of genders within the customer dataset. The x-axis represents different gender categories, while the y-axis shows the count of customers in each category.

This distribution helps Starbucks understand the gender composition of their customer base. Gender insights can assist in customizing marketing campaigns, tailoring offers, and



This scatter plot displays the relationship between **age** and **income** of the Starbucks customers. The x-axis represents **age**, while the y-axis represents **income**. The points are color-coded by **gender**, giving an additional dimension to the visualization.

This plot provides useful insights for Starbucks to understand the age-income dynamics of their customer base. These insights could help tailor product offerings or marketing campaigns based on customer demographics.

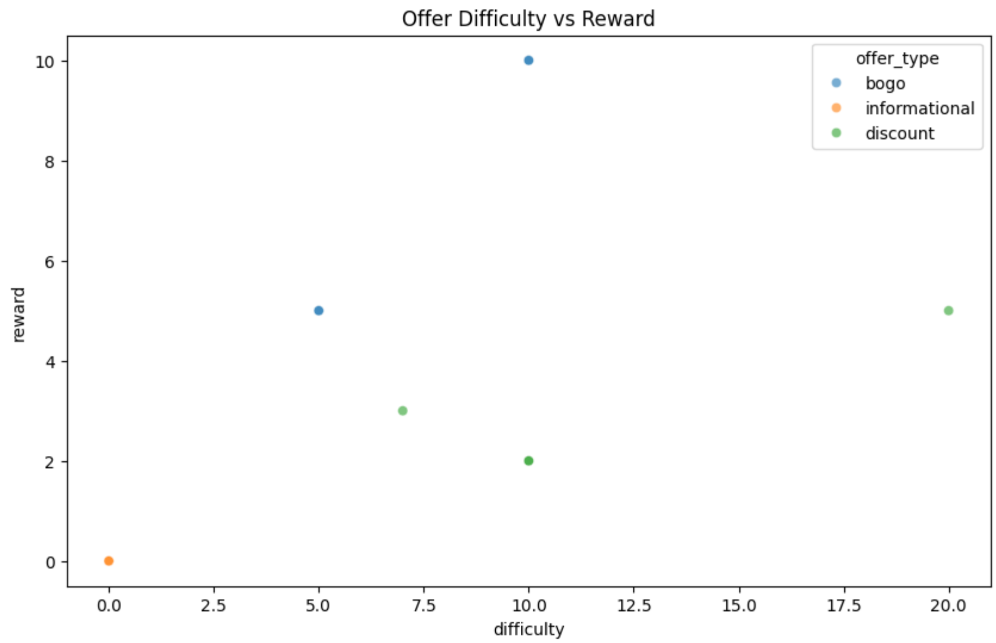


This count plot illustrates the **distribution of different offer types** provided to Starbucks customers. The x-axis represents the different **offer types** (e.g., "bogo," "discount," and "informational"), and the y-axis represents the **frequency** of each offer type.

This distribution gives Starbucks valuable insights into their promotional strategies and can be useful for analyzing customer engagement with different types of offers.

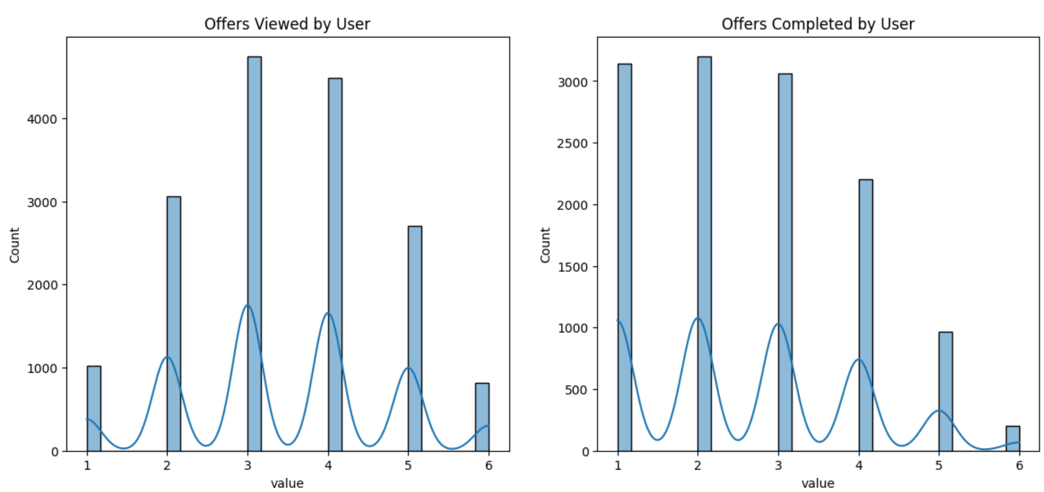
This scatter plot visualizes the **relationship between offer difficulty and reward** across different offer types. The x-axis represents the **difficulty** (the amount a customer must spend to activate the offer), and the y-axis represents the **reward** (e.g., points or benefits the customer receives). The plot is colored by **offer type** (e.g., "bogo," "discount," "informational").

This plot helps Starbucks understand how different types of offers balance difficulty and reward, which can inform future marketing strategies aimed at optimizing customer engagement.



These histograms show the **distribution of offers viewed** and **offers completed** per user. Each plot illustrates how often individual users have viewed and completed offers.

This analysis helps in identifying the customer engagement funnel, highlighting how many users progress from viewing an offer to completing it.



3. **Feature Engineering**:
    - Creating features based on customer behavior, such as the number of offers received, viewed, and completed.
    - Encoding offer details such as duration, difficulty, and reward.
4. **Modeling**:

    RandomForestRegressor: We trained this model to predict outcomes and evaluated its performance.
    DecisionTreeRegressor: This was another model used in the experiments, as mentioned in the training results

DecisionTreeRegressor_membership_duration_comparison

Understanding Membership Duration Impact:

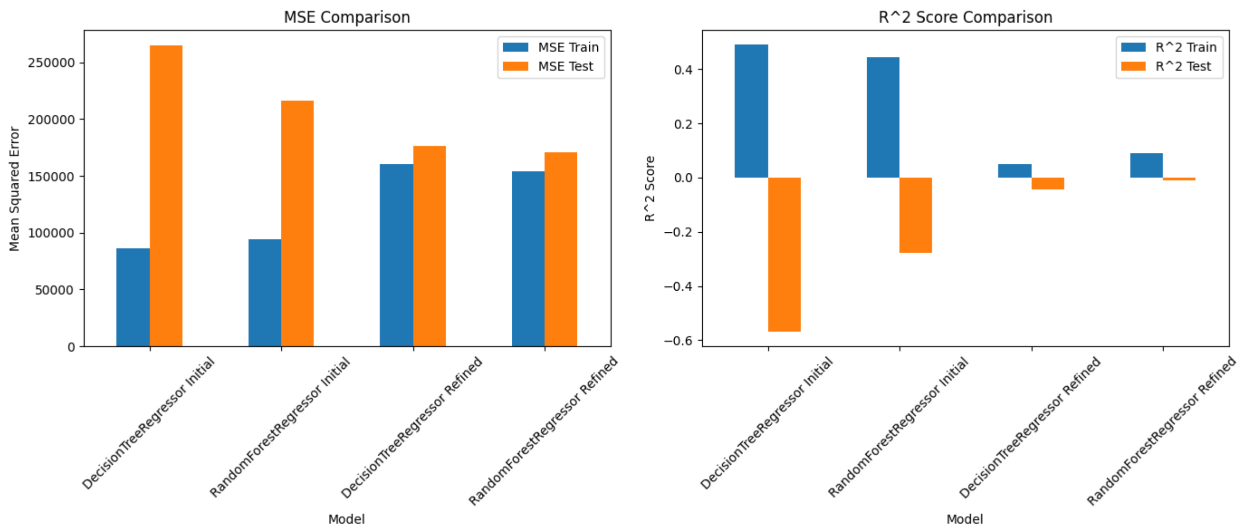
- **Model Insight:** By analyzing membership duration, the DecisionTreeRegressor can help identify how the length of a customer's membership influences their response to offers and promotions.
- **Benefit:** Starbucks can understand how long-term versus short-term members react differently to various offers, which can guide the creation of targeted promotions based on membership length.

RandomForestRegressor_membership_duration_comparison

- **Model Insight:** The RandomForestRegressor improves prediction accuracy by averaging multiple decision trees, which helps in understanding the relationship between membership duration and offer completion more reliably.
- **Benefit:** Starbucks can rely on more accurate predictions for planning and targeting offers based on membership duration.

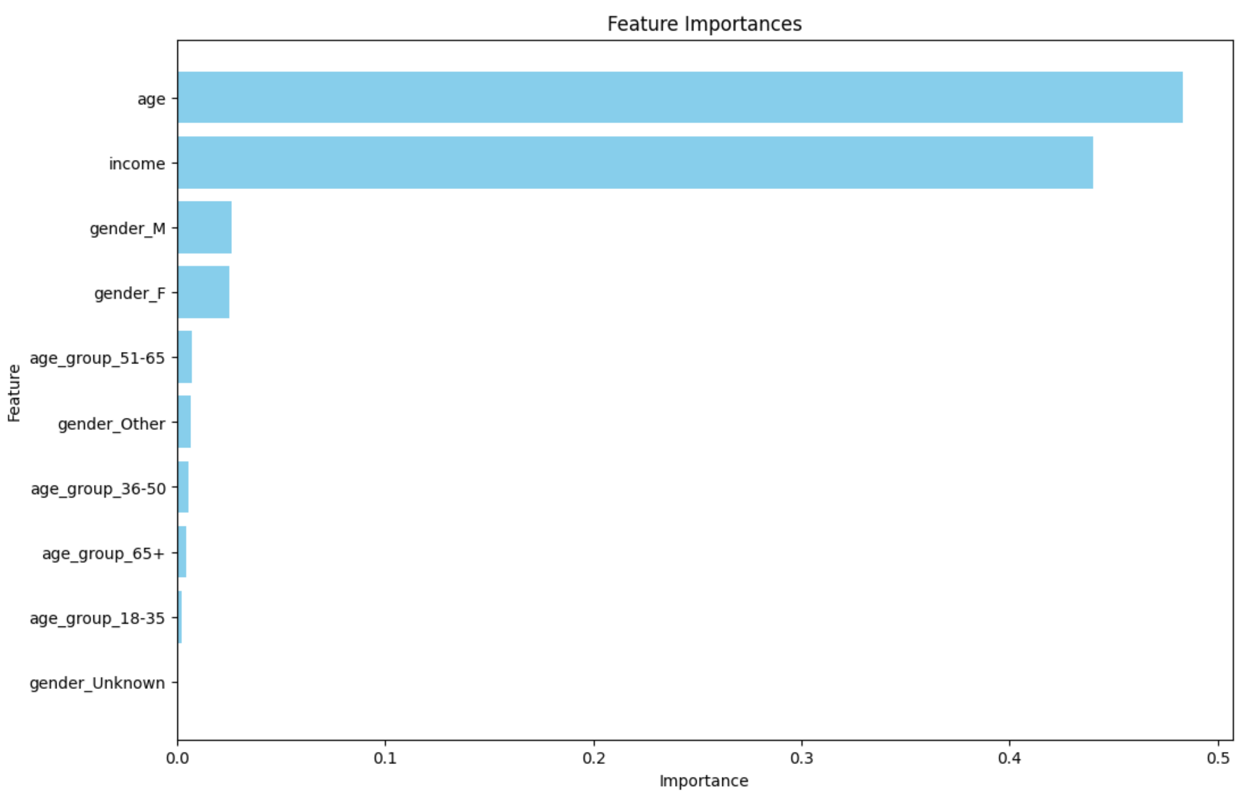
```
DecisionTreeRegressor trained on 13600 samples.
MSE_train: 86234.8584
MSE_test: 265180.3503
Training score (R^2): 0.4902
Test score (R^2): -0.5704
RandomForestRegressor trained on 13600 samples.
MSE_train: 93846.1353
MSE_test: 215540.9328
Training score (R^2): 0.4452
Test score (R^2): -0.2764
                DecisionTreeRegressor_membership_duration_comparison  \
train_time                                               0.272679
pred_time                                                0.016874
training_score                                           0.490183
testing_score                                           -0.570396

                RandomForestRegressor_membership_duration_comparison
train_time                                              21.031850
pred_time                                                0.500087
training_score                                           0.445186
testing_score                                           -0.276432
```

This plot compares the performance of the DecisionTreeRegressor and RandomForestRegressor models before and after refinement.



Feature Importances Plot

- **Purpose:** This plot visualizes the importance of each feature in the model, specifically useful for tree-based models like RandomForestRegressor.
- **Importance:** Helps in understanding which features contribute the most to the model's predictions, guiding future feature engineering and selection efforts.



How These Plots Help Starbucks

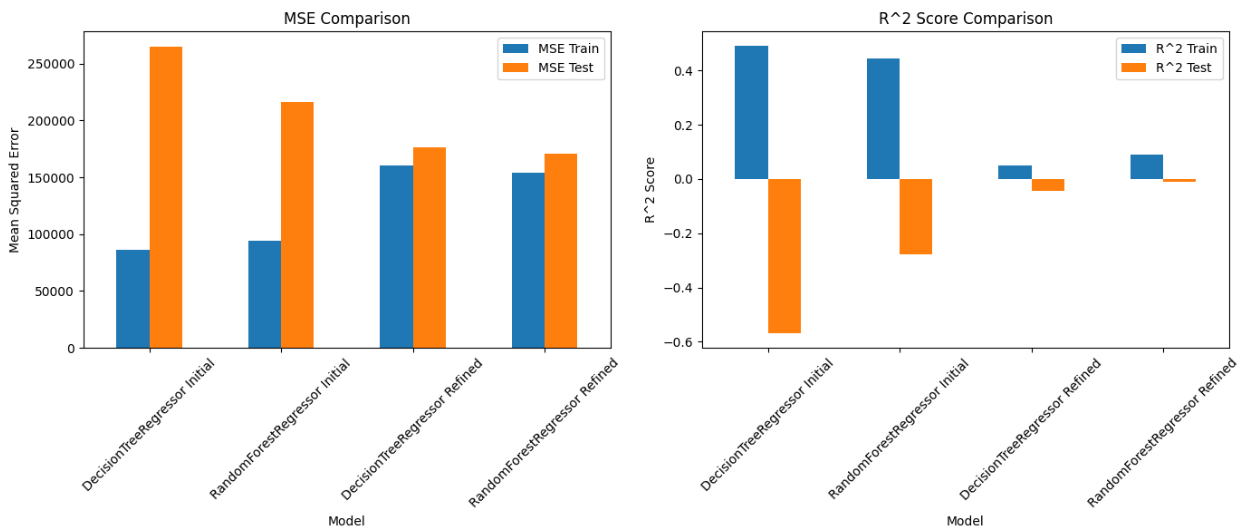
1. **Model Performance Insights:

- **Initial vs. Refined Models:** By comparing MSE and R^2 scores before and after refinement, Starbucks can assess whether model improvements are effective and if they lead to better predictions of offer completion or customer behavior.

2. **Feature Importance Understanding:

- **Actionable Insights:** Knowing which features are most important can help Starbucks focus on critical data points that affect customer behavior, improve marketing strategies, and optimize offer designs based on the most influential factors.

Model Performance The Random Forest classifier performed well in identifying customers who are likely to complete offers. By analyzing feature importance, we discovered that customer age and offer type were the most significant predictors of offer success. The Gradient Boosting model provided more nuanced predictions but required more computation time.



Conclusion This project successfully developed a machine learning model that predicts whether a customer will complete an offer. By analyzing customer demographics and transaction history, Starbucks can better tailor its marketing campaigns, delivering the right offers to the right customers. This model can potentially increase customer engagement and revenue by optimizing how offers are presented to different customer segments.

Future Work Future improvements to the model could involve incorporating additional features, such as customer app usage frequency or geographic data. Moreover, implementing real-time feedback loops could allow Starbucks to adjust offers dynamically based on immediate customer reactions.