From Data to Delivery: Heuristics for Smart Offer Deployment in the Starbucks Rewards App

Overview

In this project, I explored customer engagement with promotional offers in the Starbucks Rewards app using a **rule-based heuristic approach**. The analysis was driven by the question:

Which demographic groups are most responsive to specific types of promotional offers?

By integrating multiple data sources and applying logical heuristics, I was able to uncover actionable insights into how different customer segments engage with offers like BOGO and discounts.

Data Sources

The analysis was based on three JSON datasets from Starbucks:

- **Portfolio**: Offer metadata (types, channels, duration)
- **Profile**: Demographic data (age, gender, income, membership start date)
- **Transcript**: Customer interactions with offers (e.g., received, viewed, completed)

These datasets were cleaned and merged to form a comprehensive, analysis-ready table linking offers to user demographics and behavioral outcomes.

Motivation

From a marketing perspective, understanding how demographic groups interact with promotions enables **targeted, personalized campaigns**. This project supports that goal by revealing who responds best to which types of offers—critical information for optimizing customer retention and engagement strategies.

Problem Statement

To improve the effectiveness of Starbucks' Rewards program, I aimed to determine:

- Which demographic segments (defined by **age**, **gender**, **income**, and **tenure**) are **most responsive** to BOGO and discount offers?
- How does engagement differ when using **strict vs. flexible** definitions of offer completion?

Key Metric: Offer Completion Rate

The main performance indicator was **completion rate**, derived from user events:

- offer received
- offer viewed
- offer completed

Two definitions were used:

- **Strict Completion**: Offer must be **viewed and completed**
- **Flexible Completion**: Offer is considered complete if it was simply **completed**, regardless of viewing

---

Data Exploration & Cleaning

Portfolio Dataset

- Renamed id to offer_id
- Dropped irrelevant columns (e.g., channels, difficulty)
- Retained only offer_type and offer_id

Profile Dataset

- Dropped rows with invalid age (118) and missing gender/income (~13%)
- Converted became_member_on to datetime

- Filtered realistic age range (0–101)

Transcript Dataset

- Parsed nested dictionaries in the value column into individual fields
- Removed transaction rows unrelated to offers
- Renamed person to user_id

Data Integration

Datasets were merged using offer_id and user_id, forming a unified dataset (combined_df). Post-merge:

- Retained “transaction” events to serve as a behavioral baseline
- Removed ~11% of rows with missing demographic information for interpretability

---

Implementation Approach

The final dataset enabled **rule-based heuristics**, a logic-driven approach that does not require model training. Instead, insights were derived from **if-then rules** based on known demographic characteristics and their observed behaviors.

Example rule:

```
IF gender == 'F' AND age_group == '35-44' AND income_bracket == '>100k' THEN
recommend: BOGO offer
```

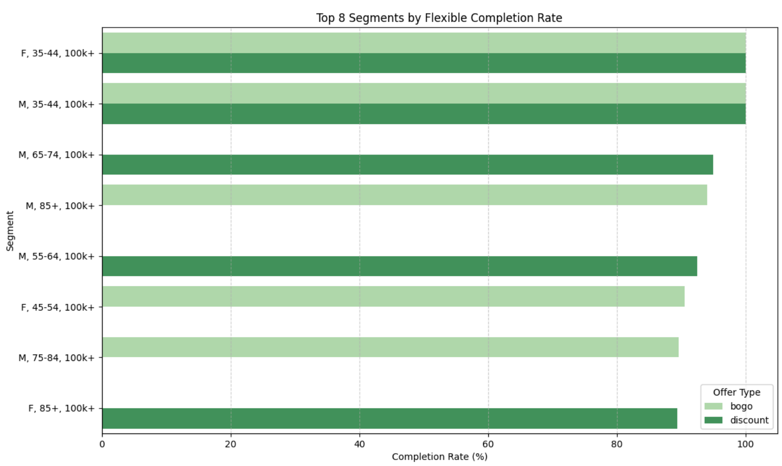
Data Visualizations & Insights

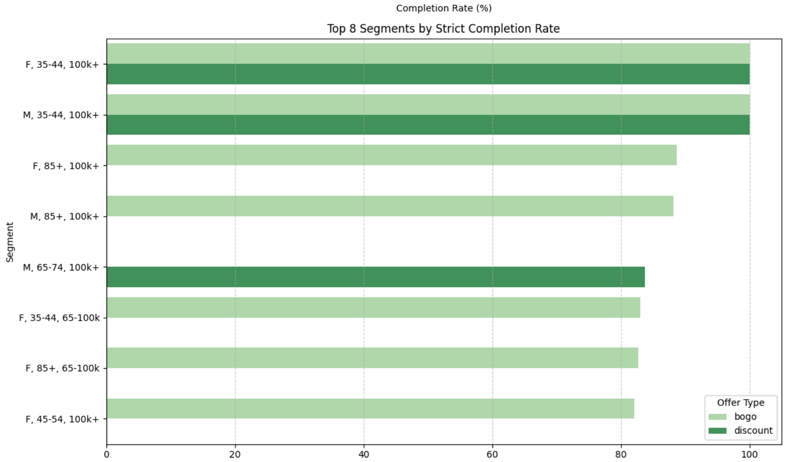
1. Bar Charts: Completion Rates by Segment

Compared **top 8 demographic segments** for each offer type under both completion definitions.

Insights:

- Flexible completion rates were consistently higher
- Females aged 35–44 earning >$100k had the highest response
- Users earning <$65k had the lowest rates, indicating a potential **marketing gap**





---

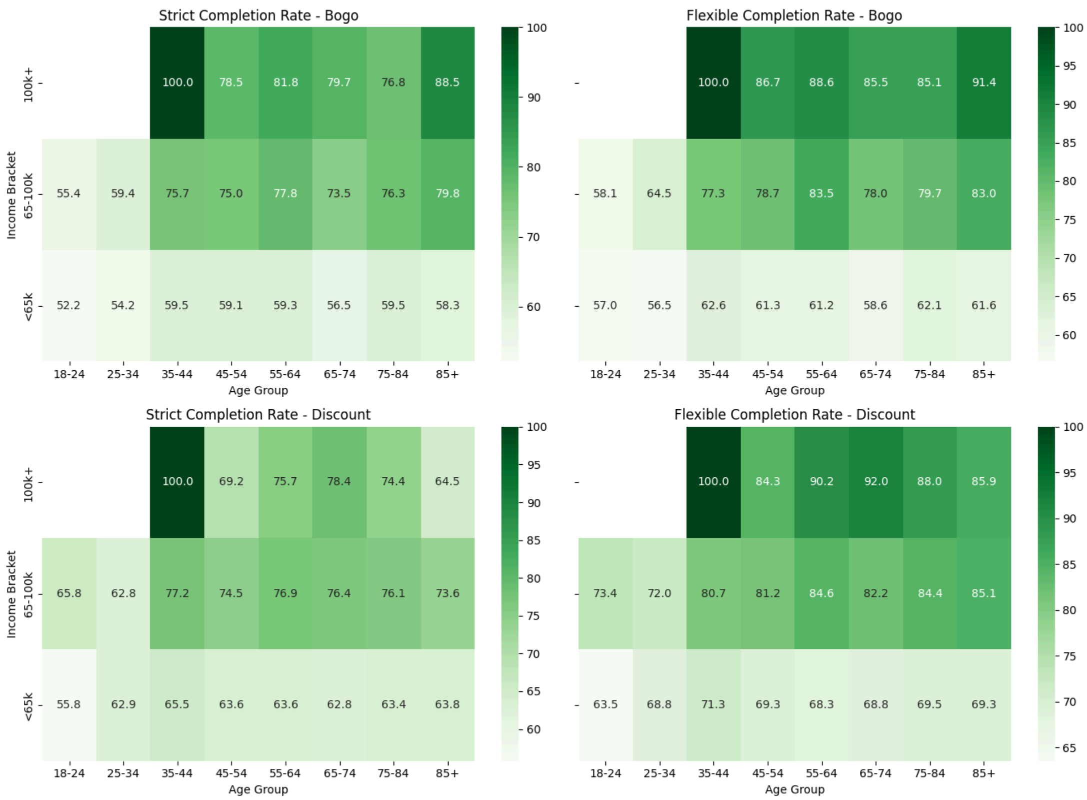
2. Heatmaps: Age vs. Income Completion Rates

A 2x2 grid visualizing:

- BOGO (strict & flexible)
  Discount (strict & flexible)

Insights:

- Completion peaks in 35–44 age group with income >$100k
- Consistently low engagement for <$65k income groups
- Business and marketing teams may prioritize different definitions:
  - **Strict** for assessing engagement
  - **Flexible** for measuring conversions



---

3. Bubble Plot: Viewed vs. Completed Offers

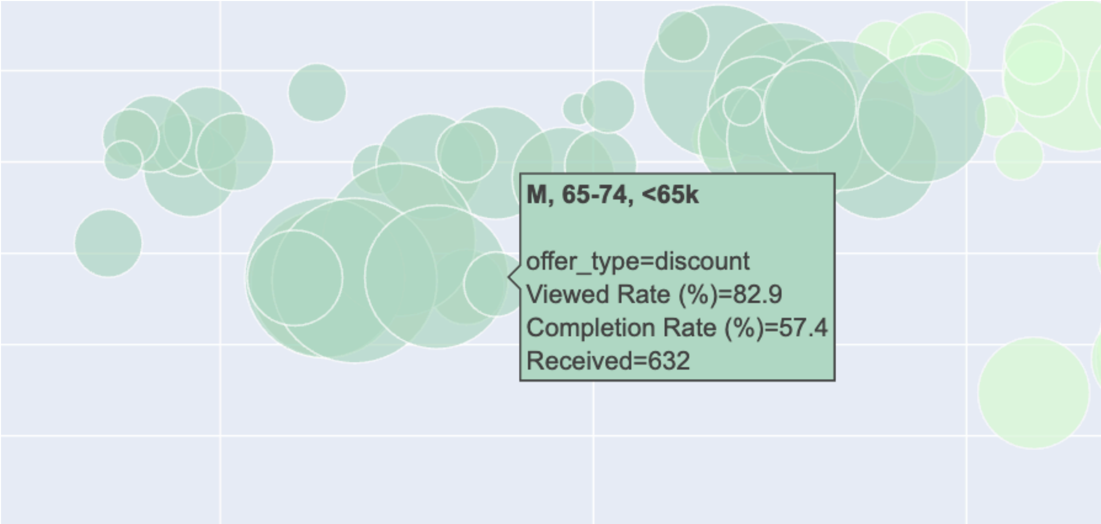
Interactive bubble plot where:

- X-axis = Viewed Rate (%)
- Y-axis = Completion Rate (%)
- Bubble Size = Number of offers received
- Color/shade = Offer type

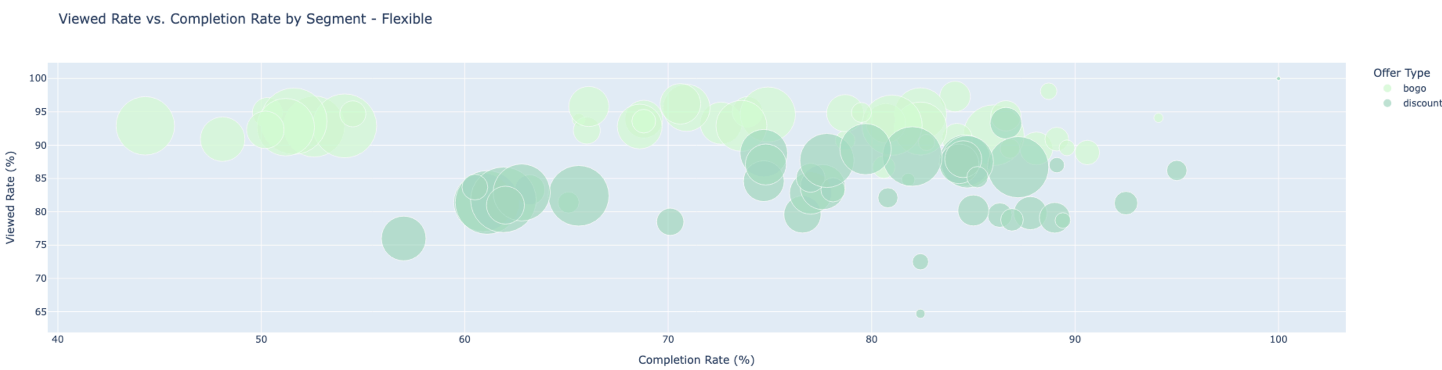
Insights:

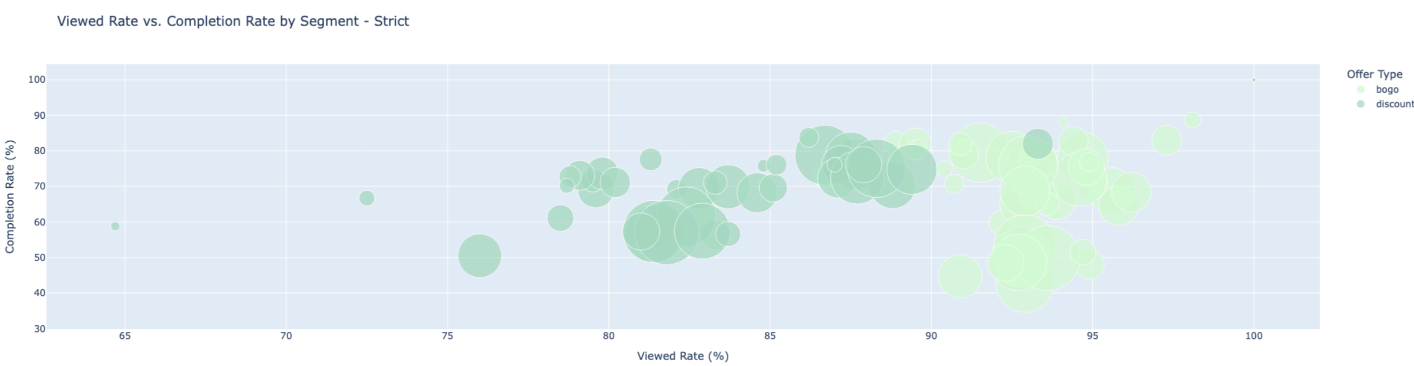
- BOGO offers had higher view rates
- Completion rates were similar across offer types

- Flexible definitions predictably resulted in higher completions



what you see when you hover over a bubble on the chart: demographic segment, offer type, viewed rate, completion rate, and number offers received in this demographic group





---

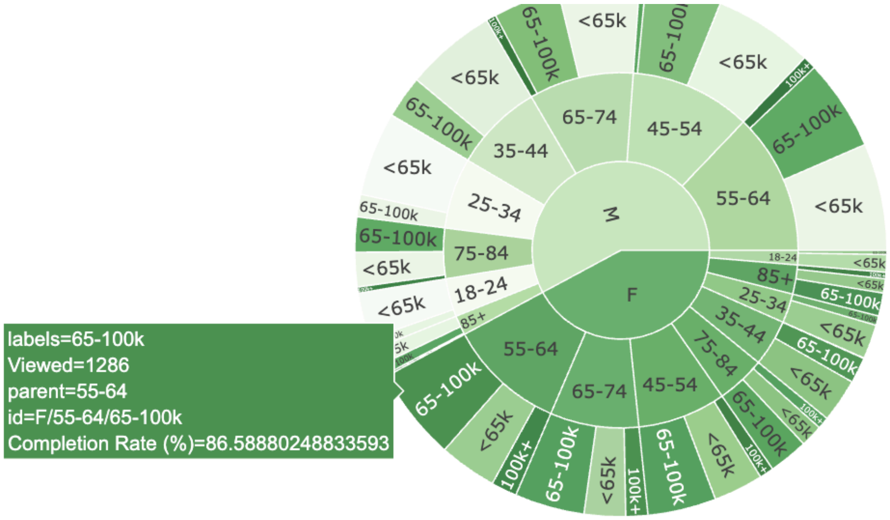
4. Sunburst Plot: Completion Hierarchy by Demographics

A hierarchical, interactive chart showing:

- Gender → Age Group → Income Bracket
- Color = Completion Rate (darker = higher)

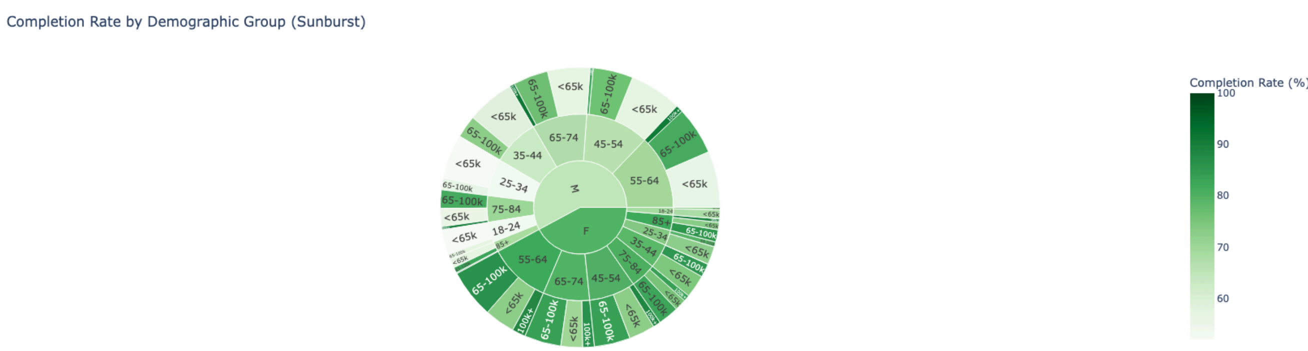
Insights:

- Female users had higher completion rates despite being fewer in number
- Visualization offered a compact way to assess both **performance** and **representation** across demographic segments



view of what you see when you hover over a slice of the sunburst

- label = income bracket of slice
- viewed = number of user in this slice that view the offer
- parent = the age group the users in this slice are in
- id = segment (gender, age group, income bracket)
- completion rate of the segment



Offer Recommendation Function

A custom function was developed to recommend an offer type based on a user's demographic segment. This logic could be integrated into the app's backend to personalize offer distribution and maximize conversions.

Conclusion

By applying a rule-based heuristic method, I was able to:

- Clean and structure messy real-world data
- Compare engagement under different definitions of success
- Visualize offer response rates across demographic groups
- Recommend offers based on user profiles

The findings suggest that **income and age** are the most predictive of responsiveness, and **personalized targeting** could dramatically improve campaign effectiveness.

Future Improvements

- Incorporate **time-based behavior analysis** (e.g., time-to-completion)
- Include **channel analysis** (email, mobile, etc.)

- Add machine learning models to complement the heuristic framework
- Improve gender inclusivity by refining the gender data handling process