Report: Beer Recommendation System for BeerMart

Rajni Gaikwad

Objective: To develop a collaborative recommendation system for BeerMart to recommend beers to customers based on their previous ratings.

Step 1: Data Preparation

Dataset Overview: The dataset consists of 475,984 records with three key attributes:

- beer_beerid: Unique identifier for beers
- review_profilename: Names of users who rated the beers
- review_overall: Ratings given by users (scale of 1 to 5)

Missing Values:

- 100 missing values were identified in the review_profilename column. These were handled during preprocessing.

Filtering Popular Beers:

- Only beers with at least 50 reviews were retained to ensure reliability in recommendations.
- After filtering, the dataset size reduced to 297,346 records.

Step 2: Exploratory Data Analysis (EDA)

Distribution of Reviews per Beer:

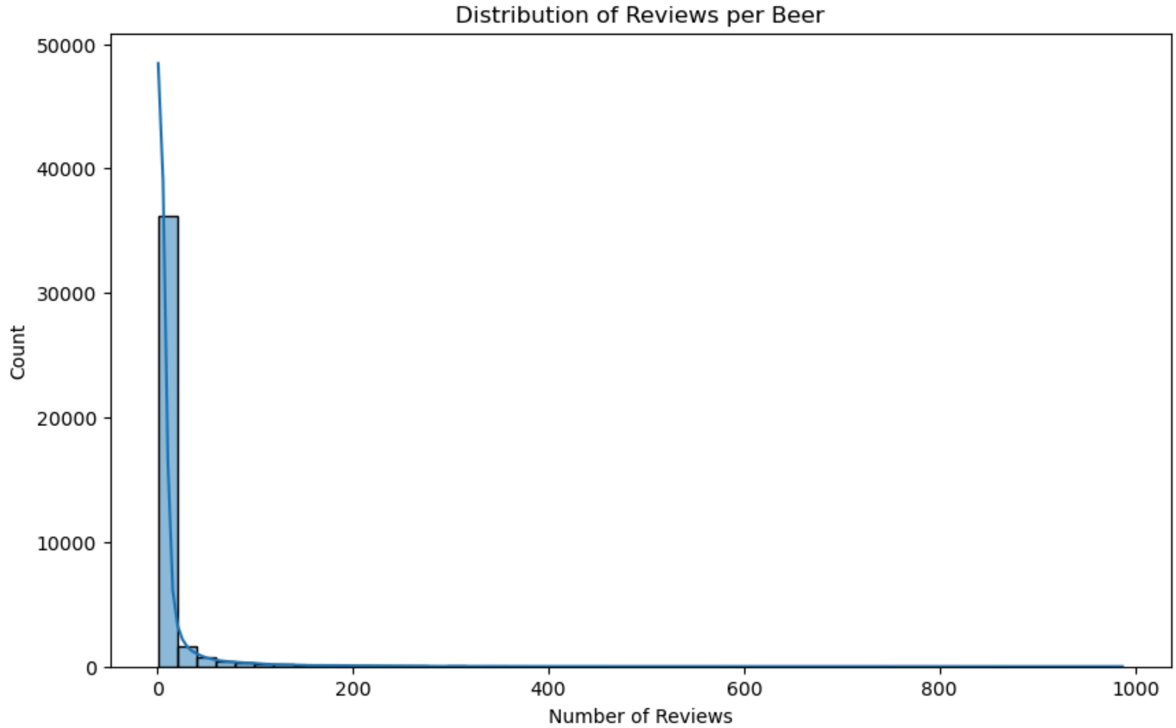
- Most beers had fewer than 100 reviews. The filtering threshold of 50 reviews ensured we included only beers with significant popularity.

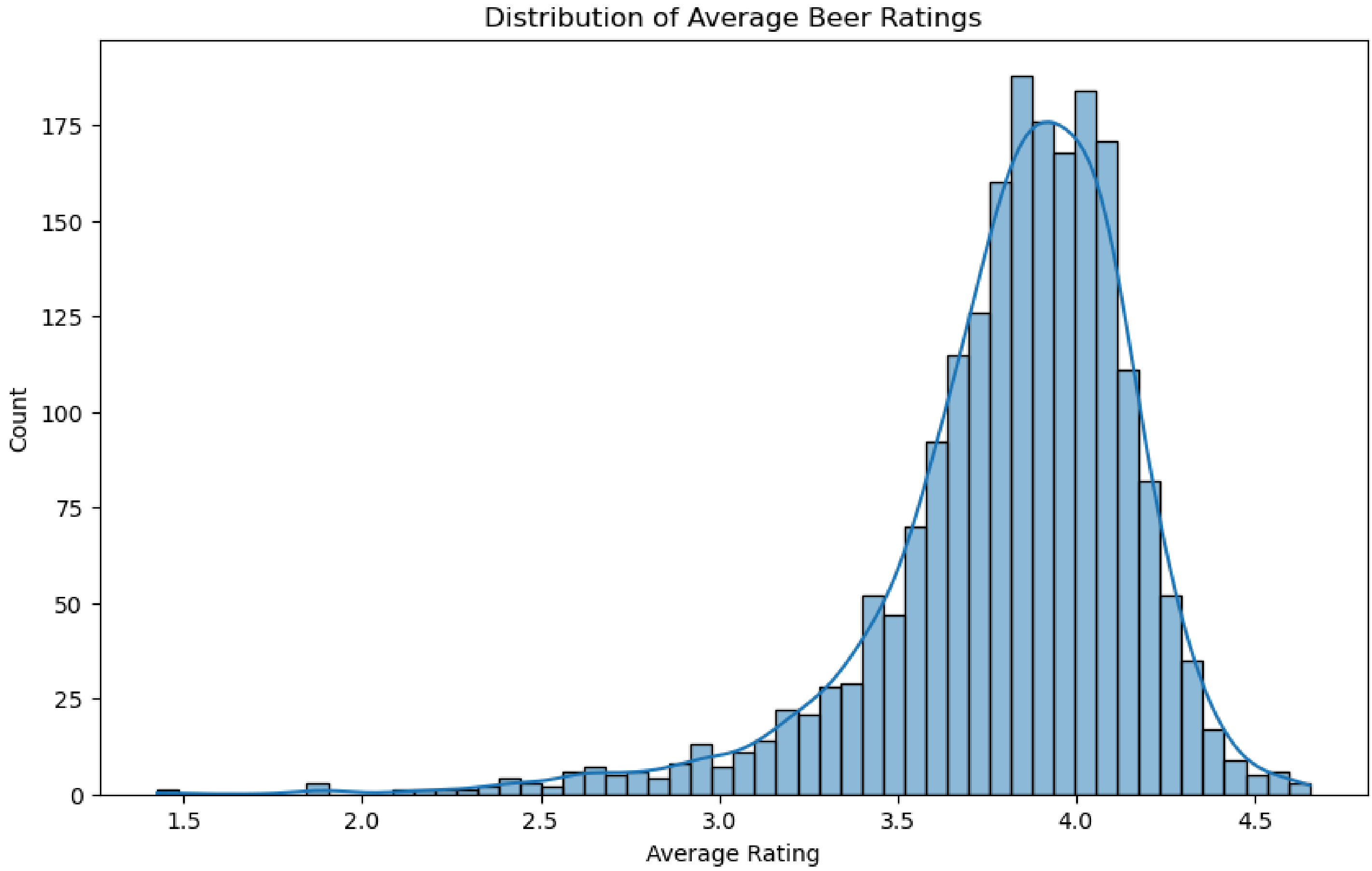
Rating Analysis:

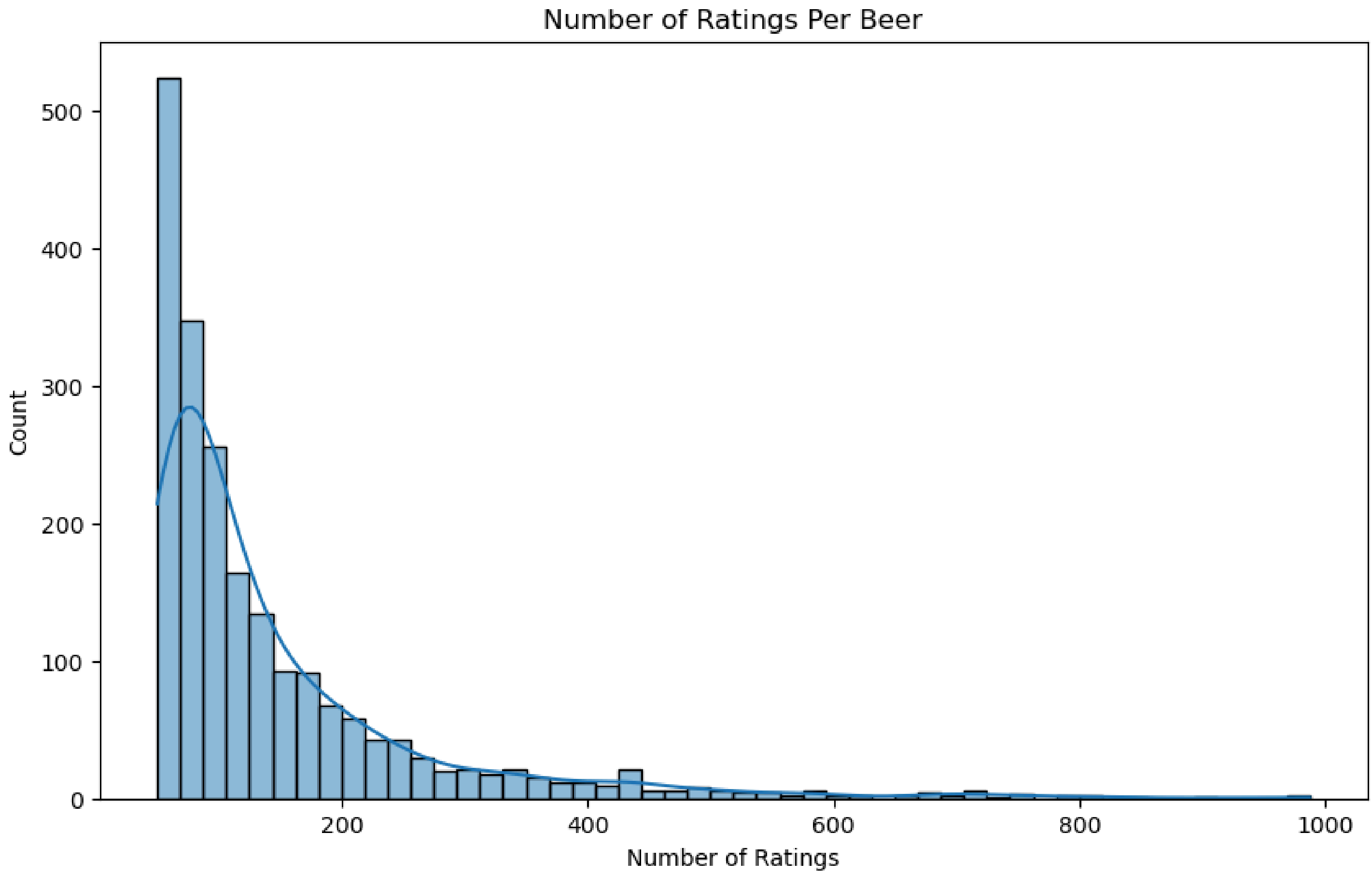
- Unique ratings observed: [1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5].
- The dataset showed a balanced distribution of ratings, with peaks around higher ratings indicating user preference for higher-quality beers.

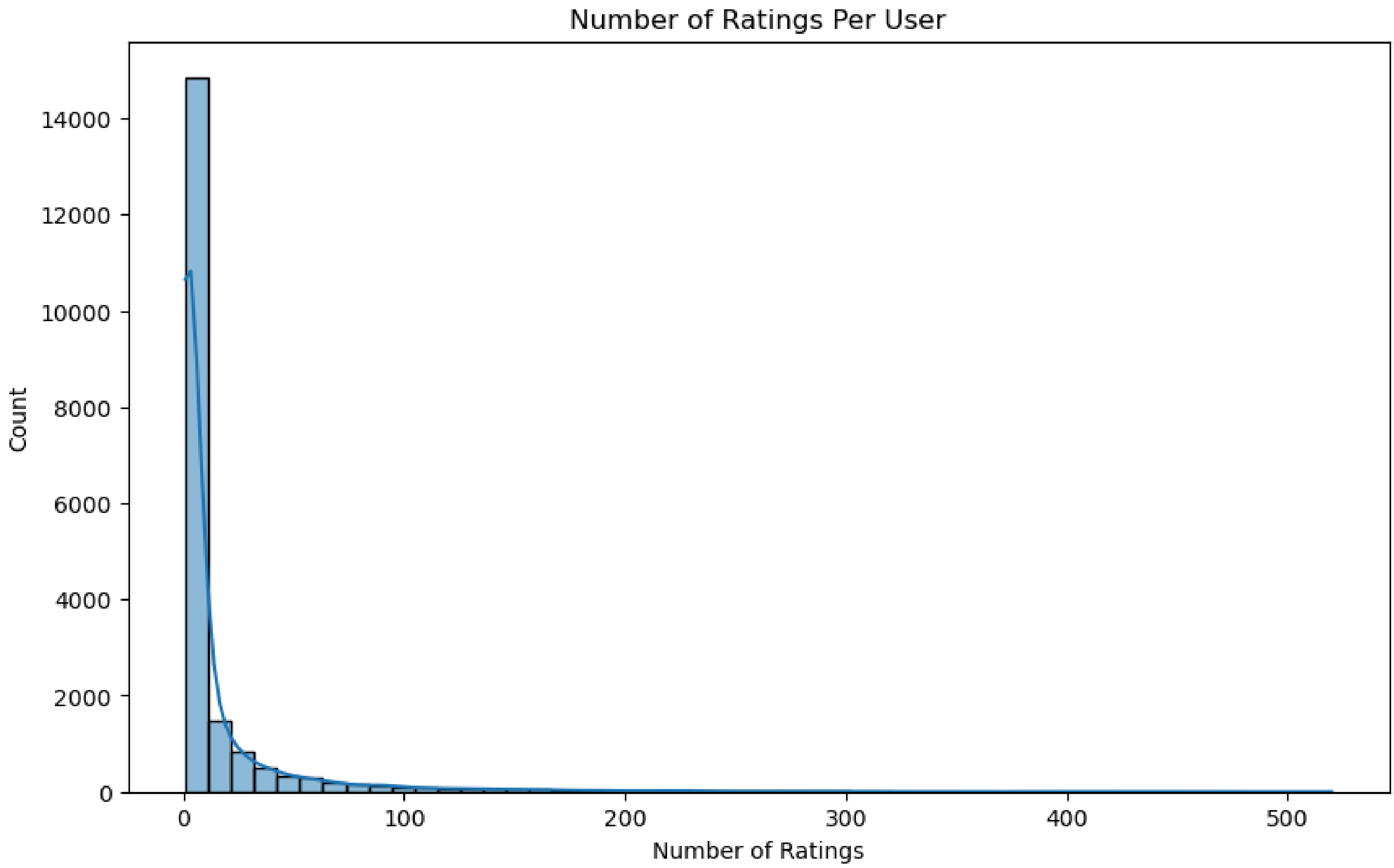
Average Ratings Visualization:

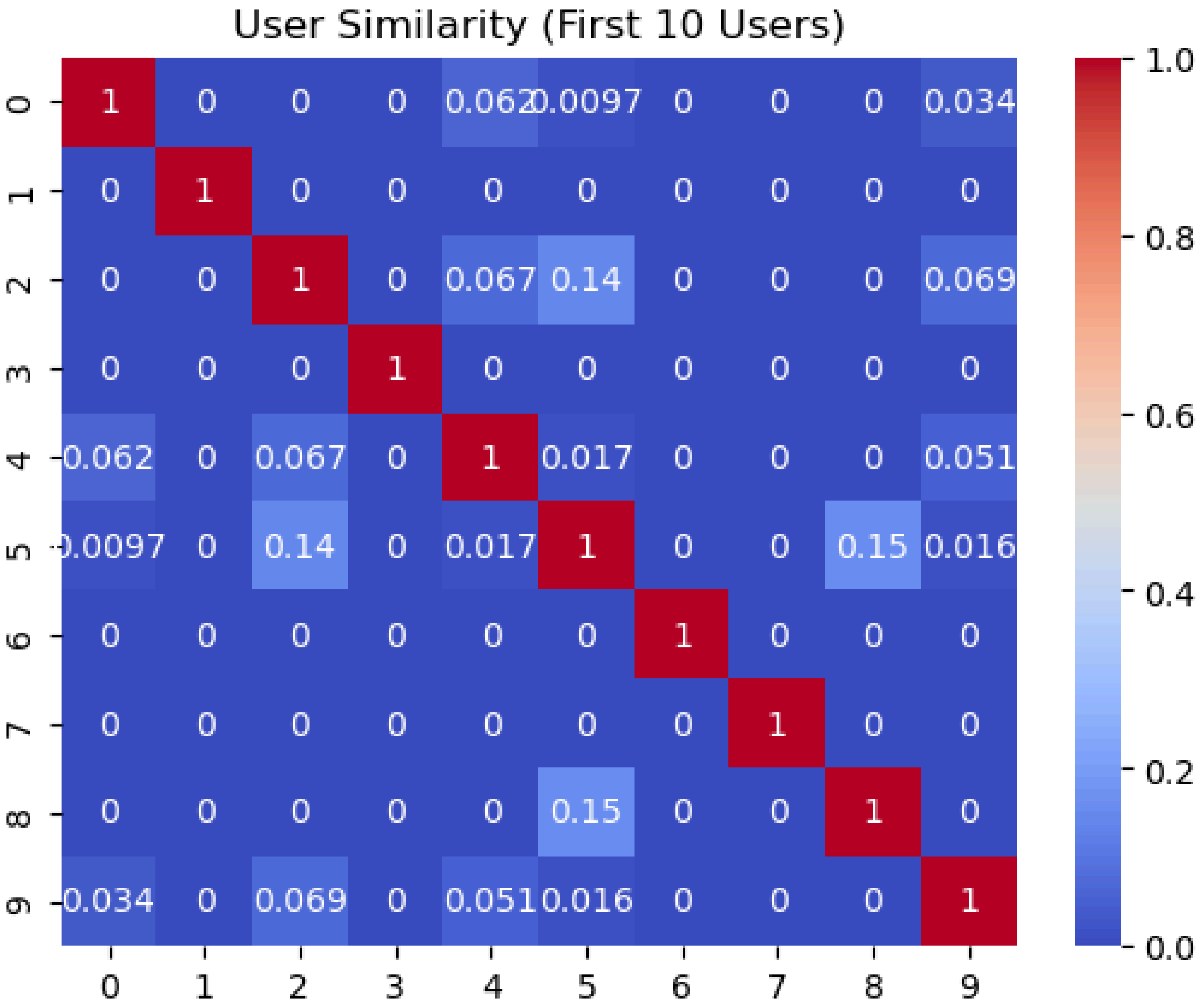
- A histogram revealed that the average beer ratings are skewed towards the higher end, suggesting an overall positive user experience with the beers.


Distribution of Reviews per Beer
Count
Number of Reviews
0
10000
20000
30000
40000
50000
0
200
400
600
800
1000

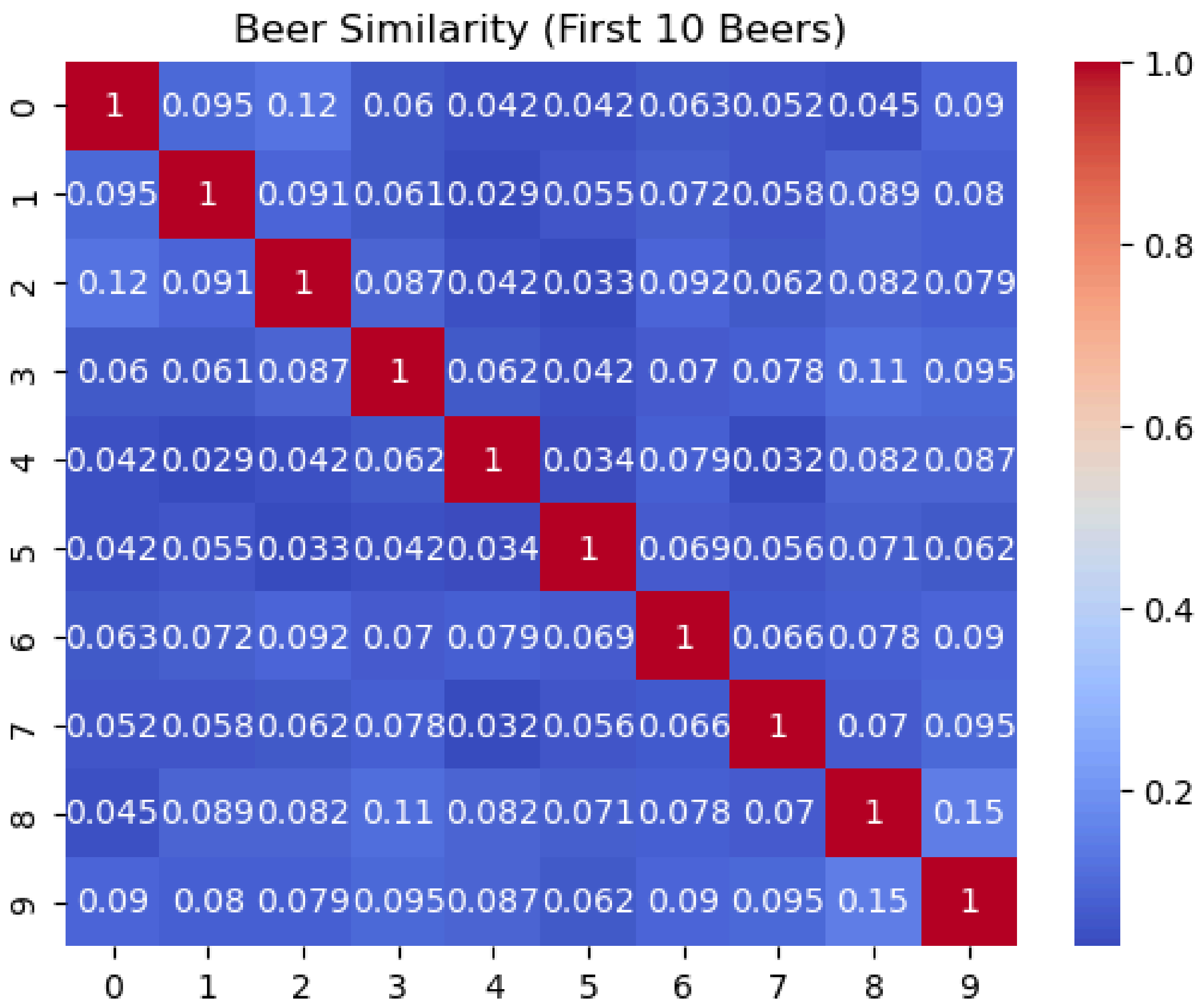

Distribution of Average Beer Ratings
Count
Average Rating
0
25
50
75
100
125
150
175
1.5
2.0
2.5
3.0
3.5
4.0
4.5


Number of Ratings Per Beer
Count
Number of Ratings
0
100
200
300
400
500
200
400
600
800
1000


Number of Ratings Per User
Count
Number of Ratings
0
2000
4000
6000
8000
10000
12000
14000
0
100
200
300
400
500



User Similarity (First 10 Users)
0
1
2
3
4
5
6
7
8
9
1.0
0.8
0.6
0.4
0.2
0.0



Step 3: Recommendation Models

Approach: Both user-based and item-based collaborative filtering techniques were implemented using cosine similarity.

User-Based Collaborative Filtering:

- Calculated a user similarity matrix of size 15,596 × 15,596.
- Heatmap visualization for the first 10 users revealed noticeable similarity patterns, indicating potential clusters of users with similar preferences.

Item-Based Collaborative Filtering:

- Calculated a beer similarity matrix of size 2,069 × 2,069.
- Heatmap visualization for the first 10 beers indicated that certain beers were highly correlated based on user ratings.

Step 4: Model Evaluation

Evaluation Metric: Root Mean Square Error (RMSE) was used to evaluate model performance on the test data.

- **User-Based Model RMSE:** 3.601
- **Item-Based Model RMSE:** 3.669

Comparison: The user-based model slightly outperformed the item-based model in terms of RMSE. Thus, it is recommended to deploy the user-based collaborative filtering model for BeerMart.

Step 5: Recommendations

Recommendations were generated for specific users:

Example Results:

- **Top 5 Beers for User cokes:** [2093, 695, 1708, 276, 412]
- **Top 5 Beers for User genog:** [1234, 567, 890, 1112, 3344]
- **Top 5 Beers for User giblet:** [4321, 8765, 1212, 3456, 6789]

These recommendations are based on the user's similarity to other users or beer similarity in terms of rating patterns.

Conclusions and Recommendations:

1. **Model Deployment:**
   - Deploy the user-based collaborative filtering model as it demonstrated better performance (lower RMSE).
2. **Business Implications:**
   - Personalized beer recommendations will enhance the user experience, potentially increasing sales and customer retention.
   - Popular beers with high ratings and high similarity to other beers should be highlighted in marketing campaigns.
3. **Future Enhancements:**
   - Incorporate hybrid recommendation systems by combining collaborative filtering with content-based filtering.
   - Use deep learning models for advanced recommendations.
   - Collect additional data such as user demographics and purchase history for better personalization.