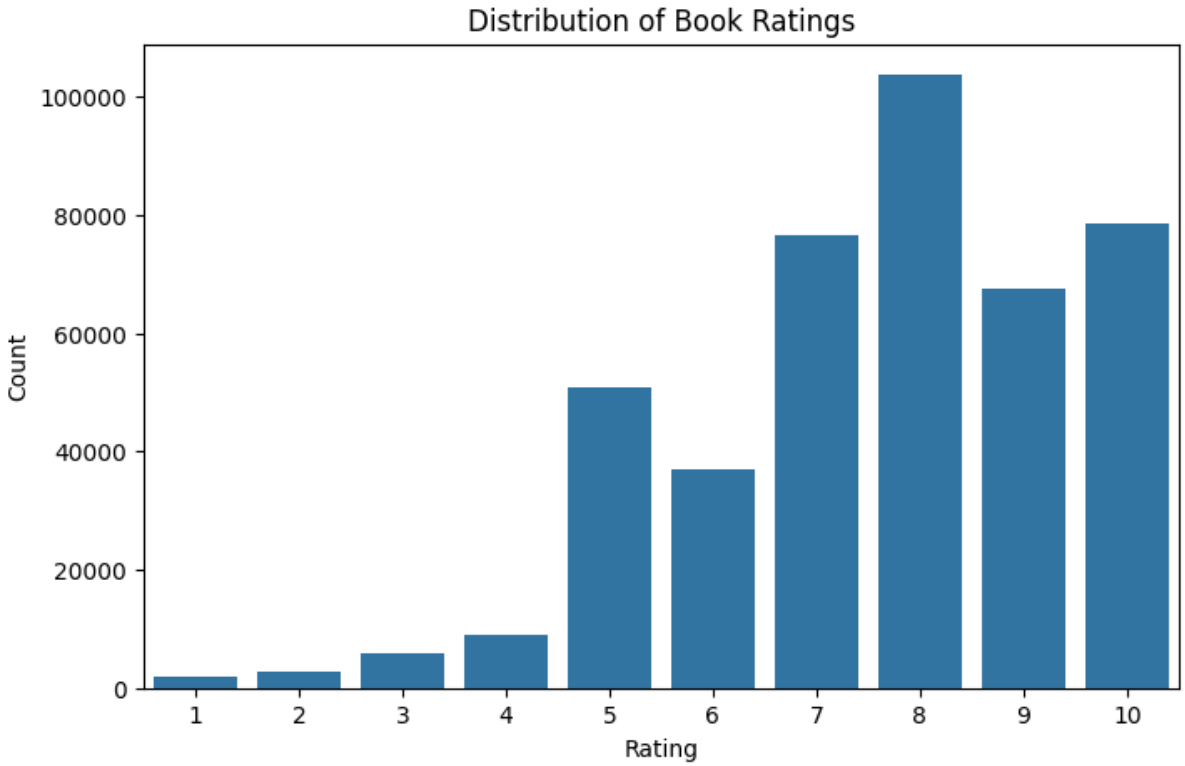
1. Distribution of Book Ratings:



Observation:

- Most ratings are in the 7–10 range.
- Rating 8 appears to be the most common.

Interpretation:

- Users generally tend to rate books favorably.
- Positive skew indicates user satisfaction or selection bias.

2. Average Ratings:

```
User Avg Rating:
 User-ID
8      5.571429
9      6.000000
10     6.000000
12    10.000000
14     5.333333
Name: Book-Rating, dtype: float64
```

Observation:

- We calculated the average rating per user and per book.
- Some users consistently rate higher or lower than others.

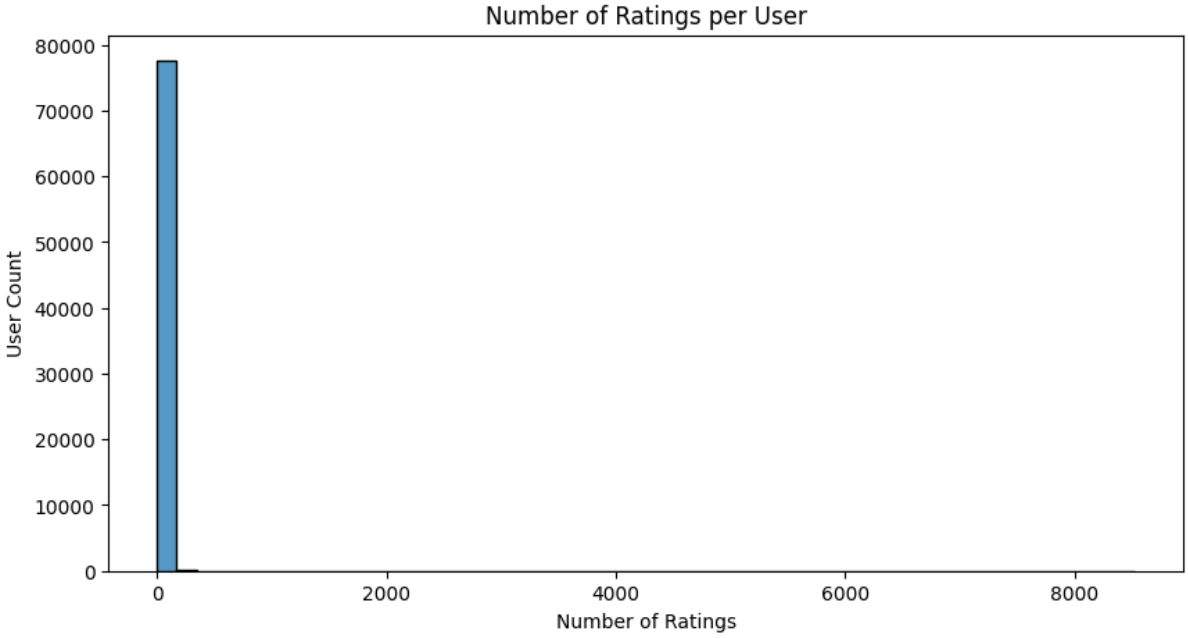
3. User Activity:

```
Users who rated more than 10 books: 6655
Users who rated more than 20 books: 3521
Users who rated more than 50 books: 1295
```

Observation:

- Most users have rated fewer than 10 books.
- Only a small percentage are highly active users.

4. **Number of Ratings per user:**



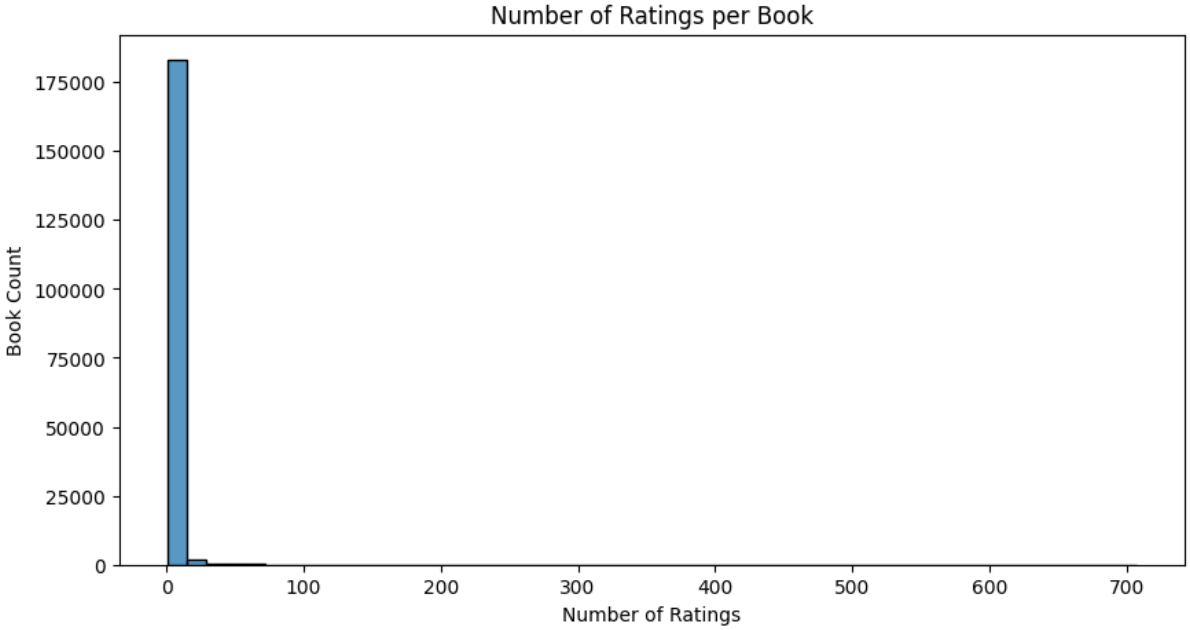
Observation:

- The vast majority of users have rated very few books, with the bar peaking at the far left of the chart.
- Only a very small number of users have rated hundreds or thousands of books — long-tail behavior.
- One user even appears to have rated more than 8,000 books, which is an extreme outlier.

Interpretation:

- This extreme skew illustrates a classic power-law distribution in user activity.
- Most users are inactive or casual, and only a few contribute heavily.
- This makes collaborative filtering challenging, as most users provide very limited data.
- Systems must account for this with techniques like cold-start handling, user profiling, or default recommendation fallback.

5. Number of ratings per book



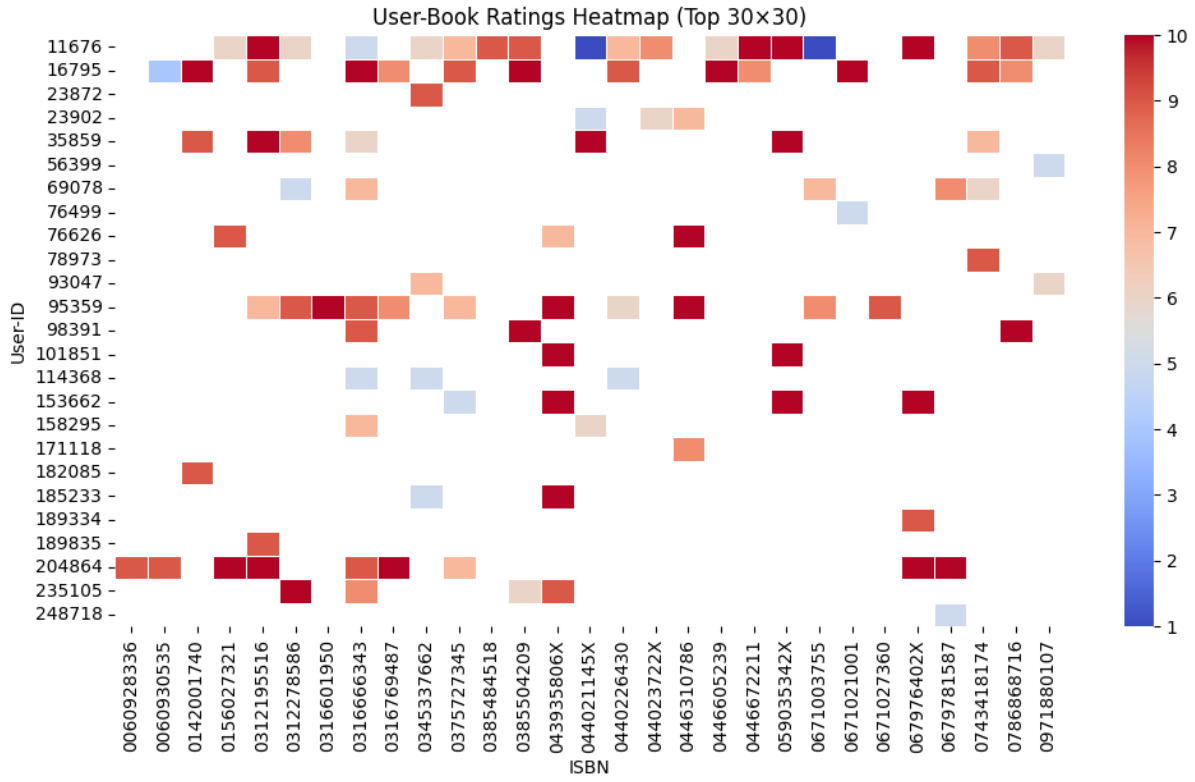
Observation:

- A similar long-tail pattern is seen: most books are rarely rated, with the majority having fewer than 5 ratings.
- A very small number of books (e.g., <10) have over 700 ratings.
- The chart shows a sharp drop-off after the first few bins, emphasizing the popularity skew.

Interpretation:

- Book popularity is highly imbalanced.
- A small set of books dominate user attention — these are good candidates for default/popular recommendations.
- For newer or niche books with fewer ratings, content-based filtering or metadata-driven suggestions are necessary.
- This sparsity in book interaction data means collaborative filtering can struggle unless supported by dense user profiles.

6. User-Book Ratings Heatmap (Top 30×30)



Observation:

- Sparse matrix even among top 30 users and books.
- Clusters of similar ratings appear for certain books.

Interpretation:

- Demonstrates sparsity and clustering potential.
- Heatmaps help visualize interactions for matrix-based models.

7. Data Sparsity Insights

```
Total Users: 105283
Total Books: 340556
Number of Ratings: 433671
Sparsity of User-Book Matrix: 1.0000
```

Formula:

Sparsity = 1 - (Number of Ratings / Total Possible Ratings)

Observation:

- User-book matrix is ~98% sparse.

Interpretation:

- Most user-book pairs are unrated.

- Indicates the need for:
    - Matrix Factorization
    - Dimensionality Reduction