📚 Book Dataset Analysis Summary

# Summary of Unique Counts

## Dataset: df\_final

👤 Unique users: 53424

📘 Unique books: 10000

🧾 Total ratings: 5976479

## Dataset: df\_books\_240

👤 Unique users: 53664

📘 Unique books: 10000

🧾 Total ratings: 6176479

## Dataset: df\_books\_528

👤 Unique users: 53952

📘 Unique books: 10000

🧾 Total ratings: 6416479

## Dataset: df\_books\_1056

👤 Unique users: 54480

📘 Unique books: 10000

🧾 Total ratings: 6856479

# Detailed Decade-wise Sanity Check (Expected vs Actual Points)

## Dataset: df\_books\_240

🧑‍🔬 Fictitious users per decade: 18

📅 2010 → 3067 books × 18 users → Expected: 55206, Found: 16095

📅 1920 → 89 books × 18 users → Expected: 1602, Found: 79078

📅 1930 → 121 books × 18 users → Expected: 2178, Found: 124115

📅 1940 → 155 books × 18 users → Expected: 2790, Found: 136872

📅 1950 → 210 books × 18 users → Expected: 3780, Found: 258874

📅 1960 → 272 books × 18 users → Expected: 4896, Found: 255159

📅 1970 → 400 books × 18 users → Expected: 7200, Found: 252194

📅 1980 → 704 books × 18 users → Expected: 12672, Found: 457448

📅 1990 → 1360 books × 18 users → Expected: 24480, Found: 875148

📅 2000 → 3121 books × 18 users → Expected: 56178, Found: 1871228

📅 2010 → 3067 books × 18 users → Expected: 55206, Found: 1313663

📅 Ancient Books → 480 books × 18 users → Expected: 8640, Found: 530243

📅 Unknown → 21 books × 18 users → Expected: 378, Found: 6362

## Dataset: df\_books\_528

🧑‍🔬 Fictitious users per decade: 31

📅 1930 → 121 books × 31 users → Expected: 3751, Found: 5324

📅 1970 → 400 books × 31 users → Expected: 12400, Found: 17600

📅 1990 → 1360 books × 31 users → Expected: 42160, Found: 23657

📅 2000 → 3121 books × 31 users → Expected: 96751, Found: 84491

📅 2010 → 3067 books × 31 users → Expected: 95077, Found: 125023

📅 1920 → 89 books × 31 users → Expected: 2759, Found: 81214

📅 1930 → 121 books × 31 users → Expected: 3751, Found: 121695

📅 1940 → 155 books × 31 users → Expected: 4805, Found: 140592

📅 1950 → 210 books × 31 users → Expected: 6510, Found: 263914

📅 1960 → 272 books × 31 users → Expected: 8432, Found: 261687

📅 1970 → 400 books × 31 users → Expected: 12400, Found: 244194

📅 1980 → 704 books × 31 users → Expected: 21824, Found: 474344

📅 1990 → 1360 books × 31 users → Expected: 42160, Found: 884131

📅 2000 → 3121 books × 31 users → Expected: 96751, Found: 1861641

📅 2010 → 3067 books × 31 users → Expected: 95077, Found: 1278343

📅 Ancient Books → 480 books × 31 users → Expected: 14880, Found: 541763

📅 Unknown → 21 books × 31 users → Expected: 651, Found: 6866

## Dataset: df\_books\_1056

🧑‍🔬 Fictitious users per decade: 66

📅 1930 → 121 books × 66 users → Expected: 7986, Found: 10648

📅 1970 → 400 books × 66 users → Expected: 26400, Found: 29681

📅 2000 → 3121 books × 66 users → Expected: 205986, Found: 221815

📅 2010 → 3067 books × 66 users → Expected: 202422, Found: 171807

📅 1920 → 89 books × 66 users → Expected: 5874, Found: 85130

📅 1930 → 121 books × 66 users → Expected: 7986, Found: 121695

📅 1940 → 155 books × 66 users → Expected: 10230, Found: 147412

📅 1950 → 210 books × 66 users → Expected: 13860, Found: 273154

📅 1960 → 272 books × 66 users → Expected: 17952, Found: 273655

📅 1970 → 400 books × 66 users → Expected: 26400, Found: 249713

📅 1980 → 704 books × 66 users → Expected: 46464, Found: 505320

📅 1990 → 1360 books × 66 users → Expected: 89760, Found: 967628

📅 2000 → 3121 books × 66 users → Expected: 205986, Found: 1861641

📅 2010 → 3067 books × 66 users → Expected: 202422, Found: 1366507

📅 Ancient Books → 480 books × 66 users → Expected: 31680, Found: 562883

📅 Unknown → 21 books × 66 users → Expected: 1386, Found: 7790

# Real vs Synthetic Ratings by Decade

## DF\_BOOKS\_240

2010 → 0 books, 6 synthetic users → real: 0, synthetic: 16095, total: 16095

1920 → 89 books, 20 synthetic users → real: 77298, synthetic: 1780, total: 79078

1930 → 121 books, 20 synthetic users → real: 121695, synthetic: 2420, total: 124115

1940 → 155 books, 20 synthetic users → real: 133772, synthetic: 3100, total: 136872

1950 → 210 books, 20 synthetic users → real: 254674, synthetic: 4200, total: 258874

1960 → 272 books, 20 synthetic users → real: 249719, synthetic: 5440, total: 255159

1970 → 400 books, 20 synthetic users → real: 244194, synthetic: 8000, total: 252194

1980 → 704 books, 20 synthetic users → real: 443368, synthetic: 14080, total: 457448

1990 → 1360 books, 20 synthetic users → real: 847948, synthetic: 27200, total: 875148

2000 → 3121 books, 20 synthetic users → real: 1808808, synthetic: 62420, total: 1871228

2010 → 3067 books, 15 synthetic users → real: 1268418, synthetic: 45245, total: 1313663

Ancient Books → 480 books, 20 synthetic users → real: 520643, synthetic: 9600, total: 530243

Unknown → 21 books, 20 synthetic users → real: 5942, synthetic: 420, total: 6362

## DF\_BOOKS\_528

1930 → 0 books, 44 synthetic users → real: 0, synthetic: 5324, total: 5324

1970 → 0 books, 44 synthetic users → real: 0, synthetic: 17600, total: 17600

1990 → 0 books, 18 synthetic users → real: 0, synthetic: 23657, total: 23657

2000 → 0 books, 28 synthetic users → real: 0, synthetic: 84491, total: 84491

2010 → 0 books, 41 synthetic users → real: 0, synthetic: 125023, total: 125023

1920 → 89 books, 44 synthetic users → real: 77298, synthetic: 3916, total: 81214

1930 → 121 books, 0 synthetic users → real: 121695, synthetic: 0, total: 121695

1940 → 155 books, 44 synthetic users → real: 133772, synthetic: 6820, total: 140592

1950 → 210 books, 44 synthetic users → real: 254674, synthetic: 9240, total: 263914

1960 → 272 books, 44 synthetic users → real: 249719, synthetic: 11968, total: 261687

1970 → 400 books, 0 synthetic users → real: 244194, synthetic: 0, total: 244194

1980 → 704 books, 44 synthetic users → real: 443368, synthetic: 30976, total: 474344

1990 → 1360 books, 27 synthetic users → real: 847948, synthetic: 36183, total: 884131

2000 → 3121 books, 17 synthetic users → real: 1808808, synthetic: 52833, total: 1861641

2010 → 3067 books, 4 synthetic users → real: 1268418, synthetic: 9925, total: 1278343

Ancient Books → 480 books, 44 synthetic users → real: 520643, synthetic: 21120, total: 541763

Unknown → 21 books, 44 synthetic users → real: 5942, synthetic: 924, total: 6866

## DF\_BOOKS\_1056

1930 → 0 books, 88 synthetic users → real: 0, synthetic: 10648, total: 10648

1970 → 0 books, 75 synthetic users → real: 0, synthetic: 29681, total: 29681

2000 → 0 books, 72 synthetic users → real: 0, synthetic: 221815, total: 221815

2010 → 0 books, 57 synthetic users → real: 0, synthetic: 171807, total: 171807

1920 → 89 books, 88 synthetic users → real: 77298, synthetic: 7832, total: 85130

1930 → 121 books, 0 synthetic users → real: 121695, synthetic: 0, total: 121695

1940 → 155 books, 88 synthetic users → real: 133772, synthetic: 13640, total: 147412

1950 → 210 books, 88 synthetic users → real: 254674, synthetic: 18480, total: 273154

1960 → 272 books, 88 synthetic users → real: 249719, synthetic: 23936, total: 273655

1970 → 400 books, 14 synthetic users → real: 244194, synthetic: 5519, total: 249713

1980 → 704 books, 88 synthetic users → real: 443368, synthetic: 61952, total: 505320

1990 → 1360 books, 88 synthetic users → real: 847948, synthetic: 119680, total: 967628

2000 → 3121 books, 17 synthetic users → real: 1808808, synthetic: 52833, total: 1861641

2010 → 3067 books, 32 synthetic users → real: 1268418, synthetic: 98089, total: 1366507

Ancient Books → 480 books, 88 synthetic users → real: 520643, synthetic: 42240, total: 562883

Unknown → 21 books, 88 synthetic users → real: 5942, synthetic: 1848, total: 7790