Capstone Project
Book Recommendation System
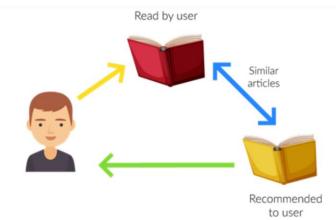
Team members
Manjusree K C

Nivya T

Ajahar

BOOK RECOMMENDATION SYSTEM

Recommendation systems are widely used today to recommend relevant products to users based on their interests. Every consumer Internet company like Netflix, YouTube etc. requires a recommendation system for efficient functioning.

A Book Recommendation System is a type of recommendation system where we have to recommend similar books to the reader based on his/her interest.

Read by user

Similar articles

Recommended to user

PROBLEM STATEMENT

Given the Book-Crossing dataset that comprises of three files-Users, Books and Ratings. The aim of this project is to build different types of book recommendation systems.

The Problem Statement

DATA SUMMARY

Books

- ISBN
- Book-Title
- Book-Author
- Year-Of-Publication
- Publisher
- Image-URL-S
- Image-URL-M
- Image-URL-L

Users

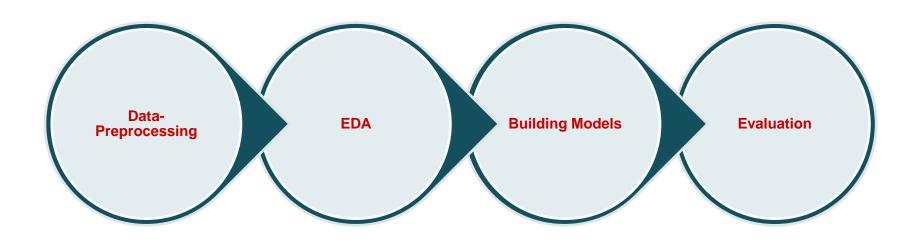- User-ID
- Location
- Age

Ratings

- User-ID
- ISBN
- Book-Rating

PRE-PROCESSING: BOOKS DATASET

- Removed the columns "Image-URL-S", "Image-URL-M" and "Image-URL-L" which do not provide any information for analysis.
- Replaced the null values in "Publisher" and "Book-Author" by "Unknown"
- Corrected misplaced entries in the columns "Book-Title", "Book-Author", "Year-Of-Publication" and "Publisher".
- Imputed invalid entries in the columns "Year of Publication" with mode.
- Cleaned "ISBN" - Converted into uppercase.
- Duplicates removal.

#	Column	Non-Null Count	Dtype
0	ISBN	271360 non-null	object
1	Book-Title	271360 non-null	object
2	Book-Author	271359 non-null	object
3	Year-Of-Publication	271360 non-null	object
4	Publisher	271358 non-null	object
5	Image-URL-S	271360 non-null	object
6	Image-URL-M	271360 non-null	object
7	Image-URL-L	271357 non-null	object

#	Column	Non-Null Count	Dtype
0	ISBN	266423 non-null	object
1	Book-Title	266424 non-null	object
2	Book-Author	266424 non-null	object
3	Year-Of-Publication	266424 non-null	int64
4	Publisher	266424 non-null	object

PRE-PROCESSING: USERS DATASET

- Cleaned Location column and separated City, state and country.
- Imputed invalid entries and null values in the column Age with median.
- Duplicates removal.

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

	User-ID	Age	City	State	Country
0	1	32	nyc	new york	usa
1	2	18	stockton	california	usa
2	3	32	moscow	yukon territory	russia
3	4	17	porto	v.n.gaia	portugal
4	5	32	farnborough	hants	united kingdom

PRE-PROCESSING: RATINGS DATASET

- Cleaned ISBN- Converted into uppercase, removed special characters.
- Removed duplicates.

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

FINAL DATASET
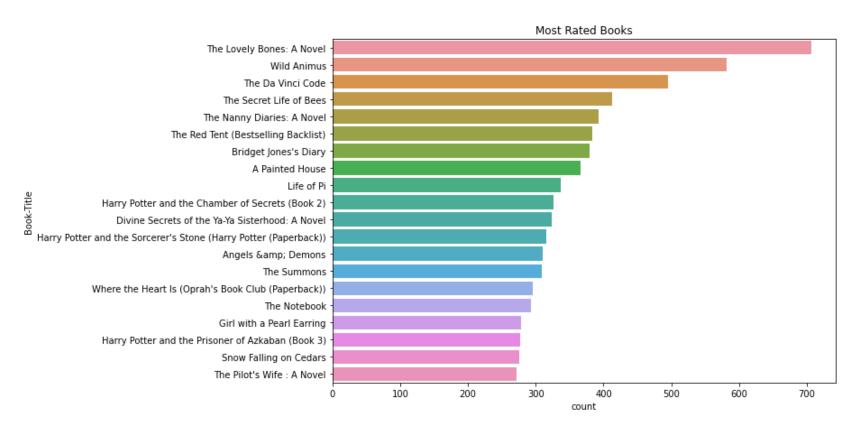

BOOKS + RATINGS + USERS = FINAL DATASET

We consider only explicit ratings for this project. So we remove 0 ratings.

```
Data columns (total 11 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   ISBN                 378257 non-null  object
 1   Book-Title           378257 non-null  object
 2   Book-Author          378257 non-null  object
 3   Year-Of-Publication  378257 non-null  int64
 4   Publisher            378257 non-null  object
 5   User-ID              378257 non-null  int64
 6   Book-Rating          378257 non-null  int64
 7   Age                  378257 non-null  int32
 8   City                 378257 non-null  object
 9   State                378257 non-null  object
 10  Country              378257 non-null  object
```
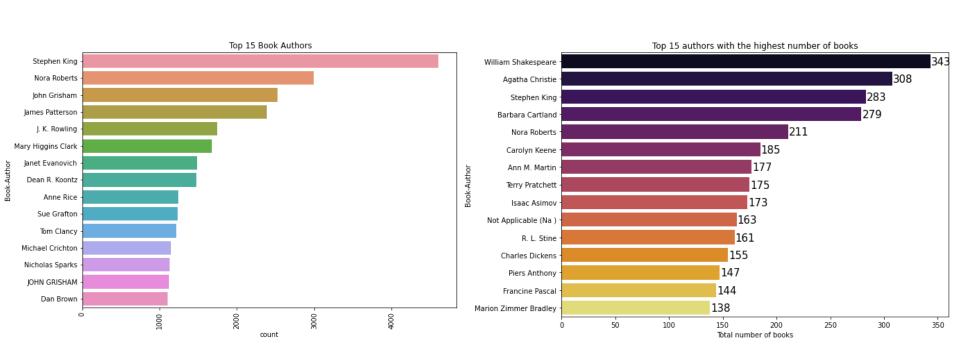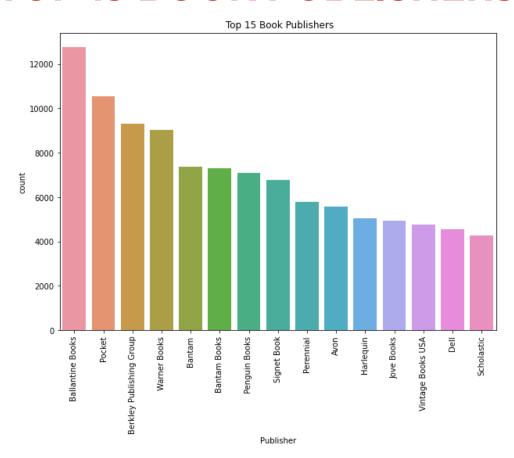
MOST RATED BOOKS

Most Rated Books

TOP BOOK AUTHORS

Top 15 Book Authors

Book-Author	count
Stephen King	
Nora Roberts	
John Grisham	
James Patterson	
J. K. Rowling	
Mary Higgins Clark	
Janet Evanovich	
Dean R. Koontz	
Anne Rice	
Sue Grafton	
Tom Clancy	
Michael Crichton	
Nicholas Sparks	
JOHN GRISHAM	
Dan Brown	

Top 15 authors with the highest number of books

Book-Author	Total number of books
William Shakespeare	343
Agatha Christie	308
Stephen King	283
Barbara Cartland	279
Nora Roberts	211
Carolyn Keene	185
Ann M. Martin	177
Terry Pratchett	175
Isaac Asimov	173
Not Applicable (Na)	163
R. L. Stine	161
Charles Dickens	155
Piers Anthony	147
Francine Pascal	144
Marion Zimmer Bradley	138

TOP 15 BOOK PUBLISHERS
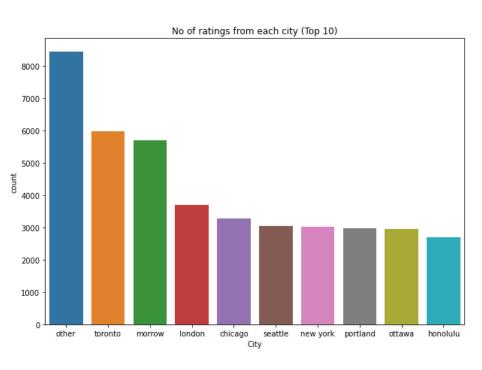

Top 15 Book Publishers

CITY AND STATE-BASED RATINGS ANALYSIS
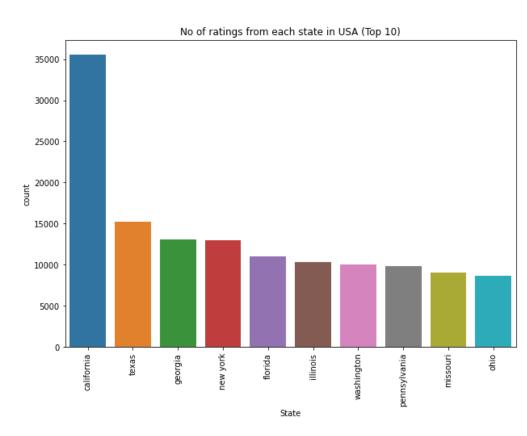

No of ratings from each city (Top 10)

No of ratings from each State (Top 10)
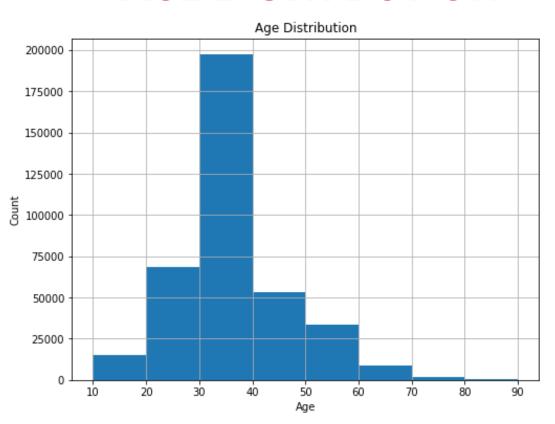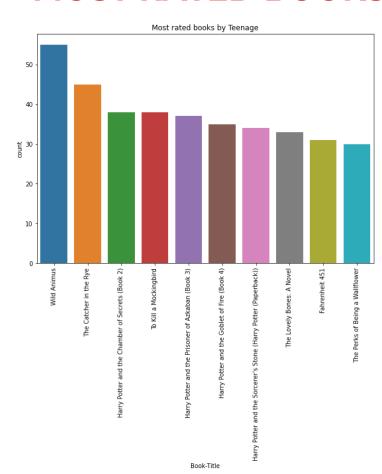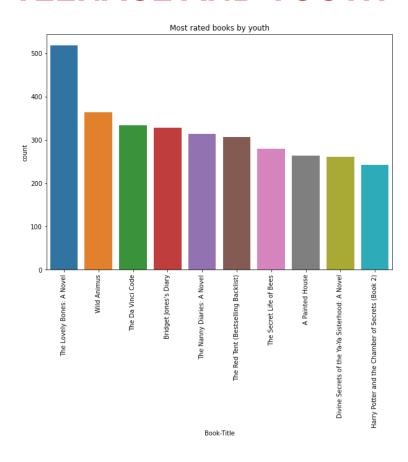
COUNTRY-BASED RATINGS ANALYSIS

AGE DISTRIBUTION

Age Distribution
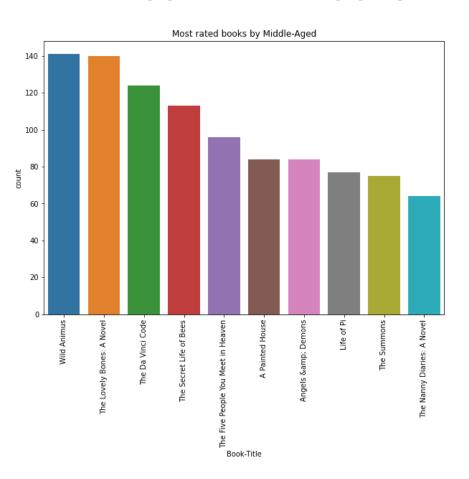
MOST RATED BOOKS BY TEENAGE AND YOUTH

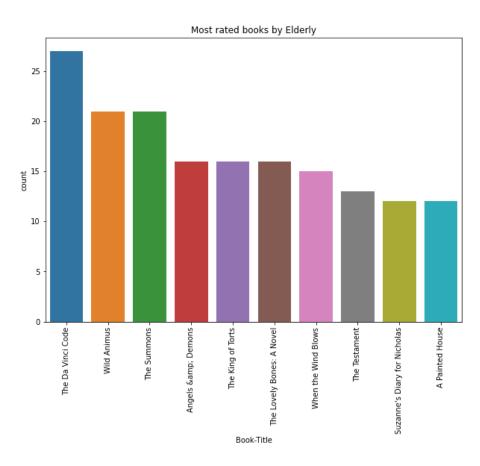# MOST RATED BOOKS BY MIDDLE-AGED AND ELDERLY

Most rated books by Middle-Aged

Most rated books by Elderly
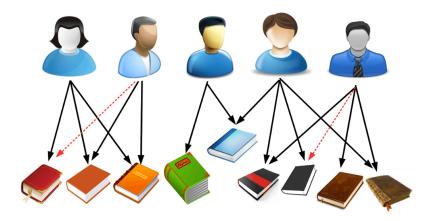
RECOMMENDATION SYSTEMS

1. Popularity-Based Recommendation System
2. Simple Recommendation System
3. Author-Based Recommendation System
4. Collaborative Recommendation - k-Nearest Neighbours
5. Collaborative Recommendation- Singular Value Decomposition

POPULARITY BASED RECOMMENDATION SYSTEM

It is a type of recommendation system which works on the principle of popularity and or anything which is in trend.

	Book-Title	Number of Ratings
0	The Lovely Bones: A Novel	707
1	Wild Animus	581
2	The Da Vinci Code	495
3	The Secret Life of Bees	413
4	The Nanny Diaries: A Novel	393
5	The Red Tent (Bestselling Backlist)	383
6	Bridget Jones's Diary	380
7	A Painted House	366
8	Life of Pi	337
9	Harry Potter and the Chamber of Secrets (Book 2)	326

SIMPLE RECOMMENDATION SYSTEM

It is a type of recommendation system which works on the principle of weighted rating.

$$\left(\frac{v}{v+m}\right) \times R + \left(\frac{m}{m+v}\right) \times c$$

	ISBN	Book-Title	Book-Author	Average Rating	Number of Votes	score
45823	0439425220	Harry Potter and the Chamber of Secrets Postcard Book	J. K. Rowling	9.869565	23	9.522671
70304	0618002235	The Two Towers (The Lord of the Rings, Part 2)	J. R. R. Tolkien	9.720000	25	9.417660
2638	0060256656	The Giving Tree	Shel Silverstein	9.750000	20	9.379672
102335	0836213319	Dilbert: A Book of Postcards	Scott Adams	9.923077	13	9.359537
131090	1888054557	Postmarked Yesteryear: 30 Rare Holiday Postcards	Pamela E. Apkarian-Russell	10.000000	11	9.340809
23973	0345339738	The Return of the King (The Lord of the Rings, Part 3)	J.R.R. TOLKIEN	9.402597	77	9.310026
102428	0836220889	Calvin and Hobbes	Bill Watterson	9.583333	24	9.289719
40331	0394800389	Fox in Socks (I Can Read It All by Myself Beginner Books)	Dr. Seuss	9.785714	14	9.284007
102382	0836218221	The Authoritative Calvin and Hobbes (Calvin and Hobbes)	Bill Watterson	9.600000	20	9.254672
86316	0743454529	My Sister's Keeper : A Novel (Picoult, Jodi)	Jodi Picoult	9.545455	22	9.235082

R = Average rating for each book

v = Number of Votes for a particular book

c = Mean vote across the whole report

m = Minimum number of votes required to be in the chart

AUTHOR-BASED RECOMMENDATION SYSTEM

Input Book: Harry Potter and the Chamber of Secrets (Book 2)
Number of Recommendations:10

```
The author of Harry Potter and the Chamber of Secrets (Book 2) is J. K. Rowling
Top books by the same author are:

Harry Potter et le prisonnier d'Azkaban
Fantastic Beasts and Where to Find Them
Quidditch Through the Ages
Harry Potter y la cÃ¡mara secreta
Harry Potter y el cÃ¡liz de fuego
Harry Potter y la piedra filosofal
Harry Potter Y El Prisionero De Azkaban
Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))
Harry Potter and the Goblet of Fire (Book 4)
Harry Potter and the Sorcerer's Stone (Book 1)
```

COLLABORATIVE RECOMMENDATION SYSTEM- K NEAREST NEIGHBOURS

AI

Input Book: Harry Potter and the Chamber of Secrets
(Book 2)
Number of Recommendations:10

Metric=Minkowski Distance

```
KNN Model Based Recommendations:

Recommended books for the book Harry Potter and the Chamber of Secrets (Book 2) are:

Harry Potter and the Prisoner of Azkaban (Book 3)
Harry Potter and the Goblet of Fire (Book 4)
Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))
Harry Potter and the Sorcerer's Stone (Book 1)
Dragons of Autumn Twilight (Dragonlance Chronicles, Vol 1)
The Shelters of Stone (Earth's Children Series, No 5)
Farmer Boy (Little House)
The Eye of the World (The Wheel of Time, Book 1)
The Secret of the Indian (Indian in the Cupboard)
The First Four Years (Little House)
```

Metric=Cosine Similarity

```
Cosine Similarity  Based Recommendations:

Recommended books for the book Harry Potter and the Chamber of Secrets (Book 2) are:

Harry Potter and the Prisoner of Azkaban (Book 3)
Harry Potter and the Goblet of Fire (Book 4)
Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))
Harry Potter and the Sorcerer's Stone (Book 1)
Harry Potter and the Order of the Phoenix (Book 5)
The Fellowship of the Ring (The Lord of the Rings, Part 1)
Charlotte's Web (Trophy Newbery)
Dragons of Autumn Twilight (Dragonlance Chronicles, Vol 1)
Farmer Boy (Little House)
The Eye of the World (The Wheel of Time, Book 1)
```

COLLABORATIVE FILTERING-SINGULAR VALUE DECOMPOSITION

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher
0	0743418174	Good in Bed	Jennifer Weiner	2002	Washington Square Press
1	0671027360	Angels & Demons	Dan Brown	2001	Pocket Star
2	0525947299	No Second Chance	Harlan Coben	2003	Dutton Books
3	0385510438	The Last Juror	John Grisham	2004	Doubleday
4	0375412530	Back When We Were Grownups: A Novel	ANNE TYLER	2001	Knopf
5	0786868716	The Five People You Meet in Heaven	Mitch Albom	2003	Hyperion
6	0399149155	Q Is for Quarry	Sue Grafton	2002	Putnam Publishing Group
7	0151008116	Life of Pi	Yann Martel	2002	Harcourt
8	0385335482	Confessions of a Shopaholic (Summer Display Opportunity)	SOPHIE KINSELLA	2001	Delta
9	0316603287	The Lake House	James Patterson	2003	Little, Brown

EVALUATION

```
Global metrics:
{'modelName': 'Collaborative Filtering', 'recall@5': 0.30250839074368485, 'recall@10': 0.4128687511040452}
```

	hits@5_count	hits@10_count	interacted_count	recall@5	recall@10	_person_id
2	29	80	536	0.054104	0.149254	11676
94	50	67	138	0.362319	0.485507	98391
146	48	51	92	0.521739	0.554348	153662
290	20	39	88	0.227273	0.443182	16795
188	46	55	72	0.638889	0.763889	114368
112	19	19	72	0.263889	0.263889	95359
320	25	30	61	0.409836	0.491803	104636
163	20	24	54	0.370370	0.444444	158295
741	25	40	54	0.462963	0.740741	123883
135	5	15	52	0.096154	0.288462	204864

CONCLUSION

- The most rated books are **The Lovely Bones: A Novel** and **Wild Animus.**
- Top book author with respect to the number of ratings is **Stephan King** and with respect to the number of books written is **William Shakespeare.**
- Top publishers on the basis of the number of ratings received for their books are **Ballantine Books** and **Pocket**.
- Majority of the readers are of the age bracket 30–40 and most of them are from the countries **USA**, **Texas** and **Georgia**.
- **Wild animus** is one of the most rated books irrespective of the age.
- We built 5 types of recommendation systems and did evaluation for one of them.
- After evaluation for Collaborative-Model Based Recommendation system, we got a recall@5 of 30% and recall@10 of 41%.

THANK YOU