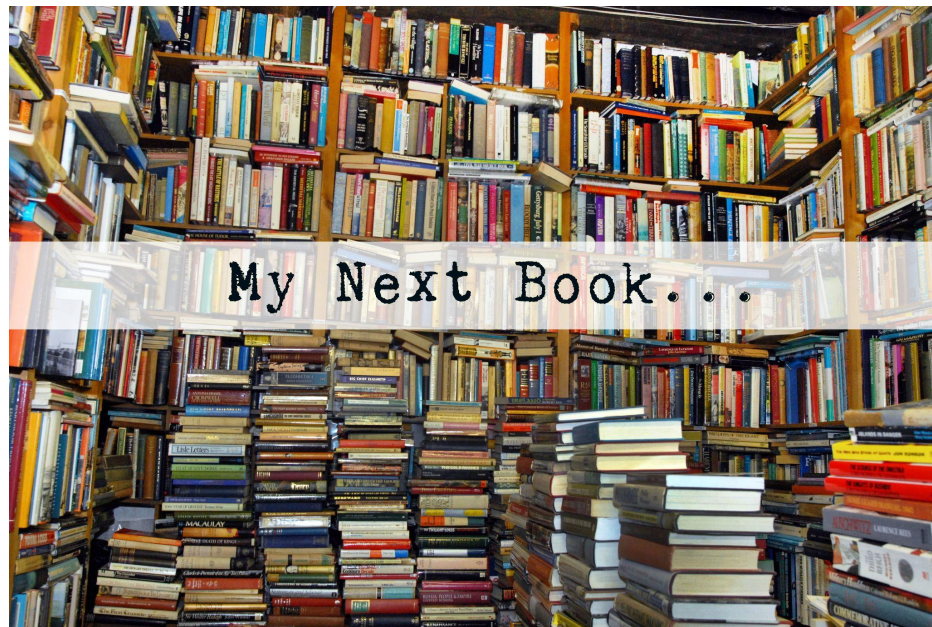


Book Recommendation System

Submitted By:
Lakshmi Narayana
Data Science Trainee
At Alma Better

Book Recommendation System



- Problem Statement
- EDA and Feature engineering
- Data preparation
- Model Implementation
- Evaluation of Model

Problem Statement



During the last few decades, with the rise of YouTube, Amazon, Netflix, and many other such web services, recommender systems have taken more and more place in our lives. From e-commerce (suggest to buyers articles that could interest them) to online advertisement (suggest to users the right contents, matching their preferences), recommender systems are today unavoidable in our daily online journeys.

Data Pipeline

- **Data processing-1:** Checked for null values in Users dataset and cleaned outliers.
- **EDA:** Exploratory analysis was done on the features selected in the first phase.
- **Data processing-2:** Checked for null values in Books dataset, cleaned outliers and changed the datatype of features.
- **EDA:** Exploratory analysis was done on the features selected in the above phase.
- **Data processing-3:** Cleaned null values in Ratings dataset.
- **EDA:** Combined the all datasets and made eda on the above selected features.
- **Create a model:** Finally in this part created models, trained and tested it on the available dataset.



Data Description

Attribute Information:

Users

Contains the users. Note that user IDs (User-ID) have been anonymised and map to integers. Demographic data is provided (Location, Age) if available. Otherwise, these fields contain NULL values.

Books

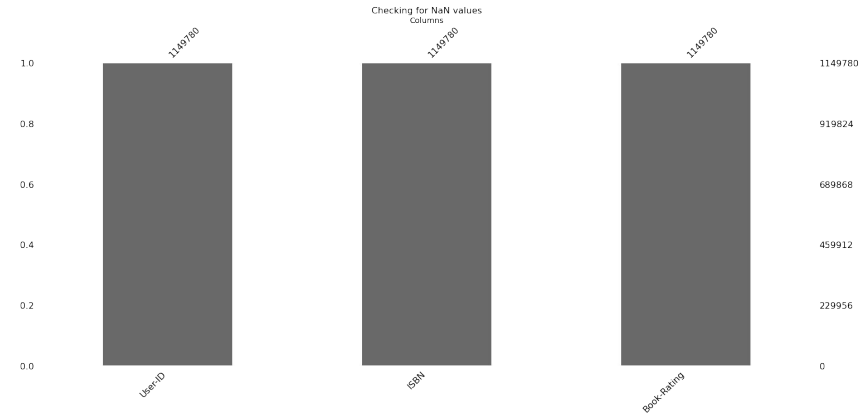
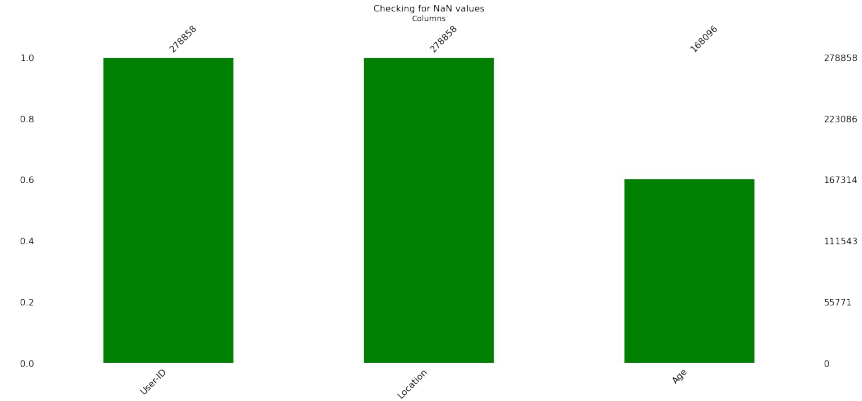
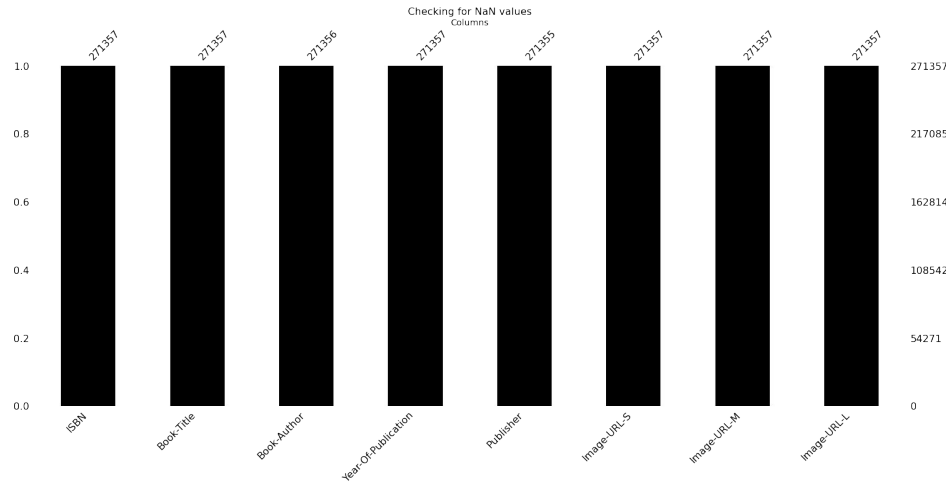
Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (Book-Title, Book-Author, Year-Of-Publication, Publisher), obtained from Amazon Web Services. Note that in the case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (Image-URL-S, Image-URL-M, Image-URL-L), i.e., small, medium, large. These URLs point to the Amazon website.

Ratings

Contains the book rating information. Ratings (Book-Rating) are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.

Exploratory Data Analysis

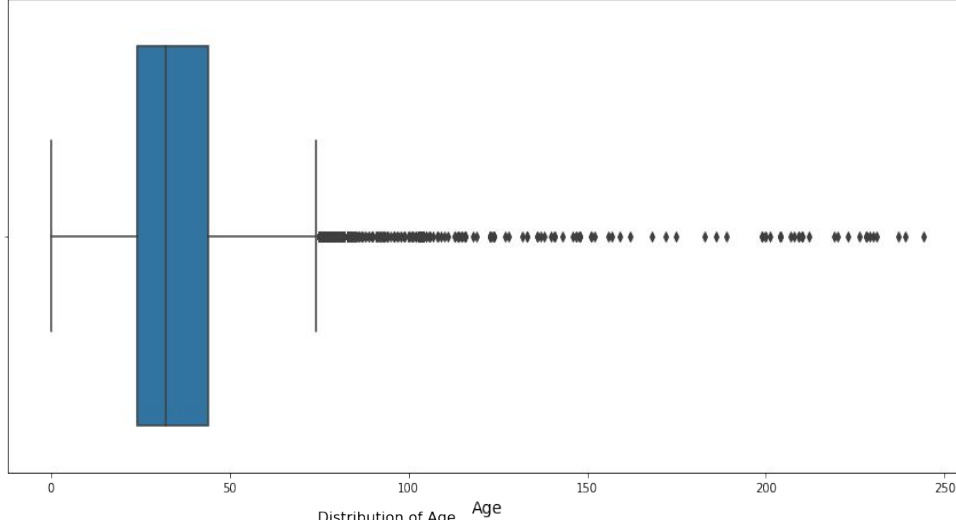
- Cleaning the null values.
- Separating columns.
- Replaced outliers.
- Exploratory Analysis.



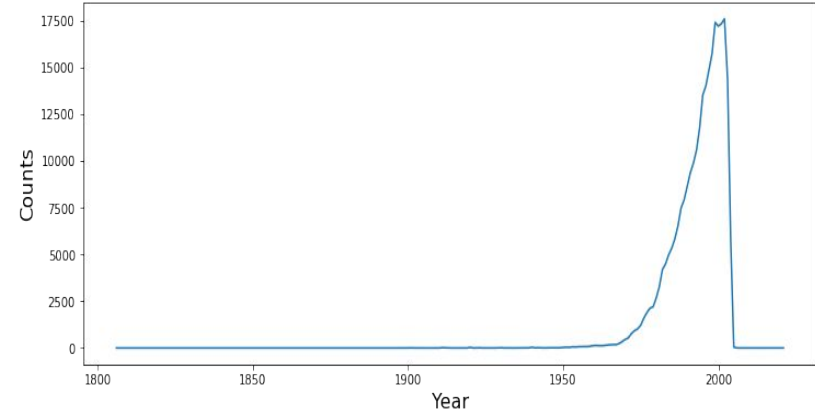
Exploratory Data Analysis



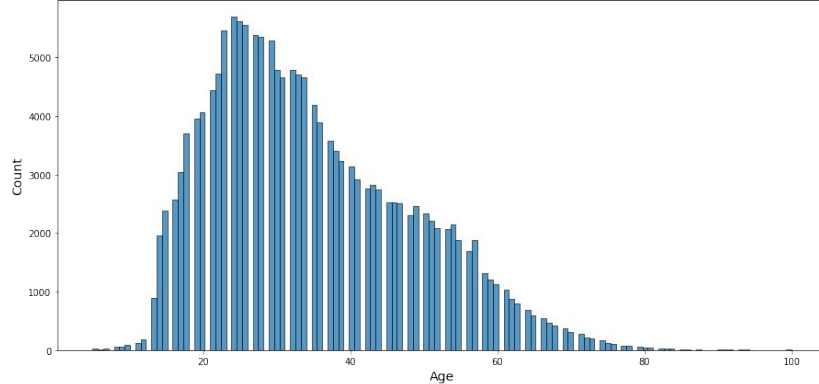
Checking for Outliers



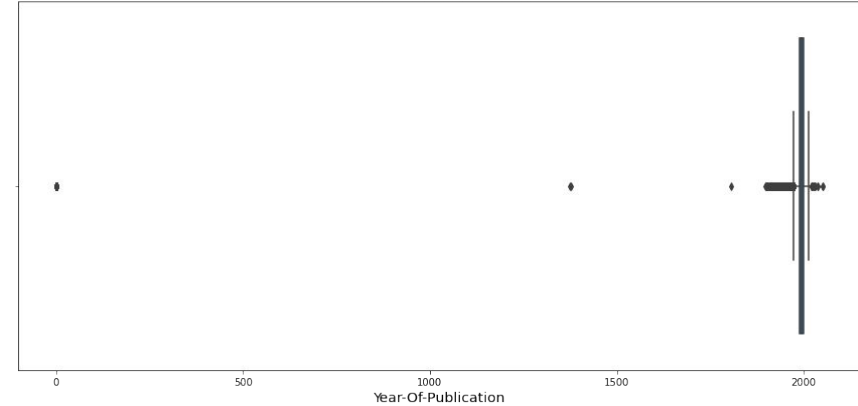
Publication Of Books Year wise



Distribution of Age

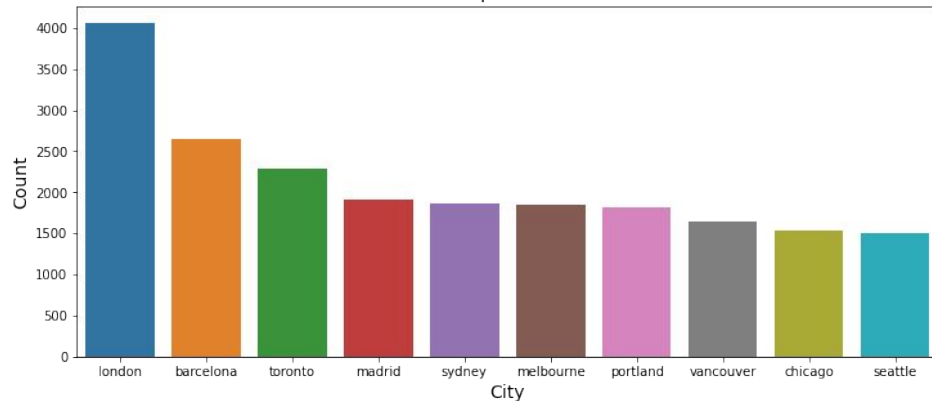


Checking for Outliers

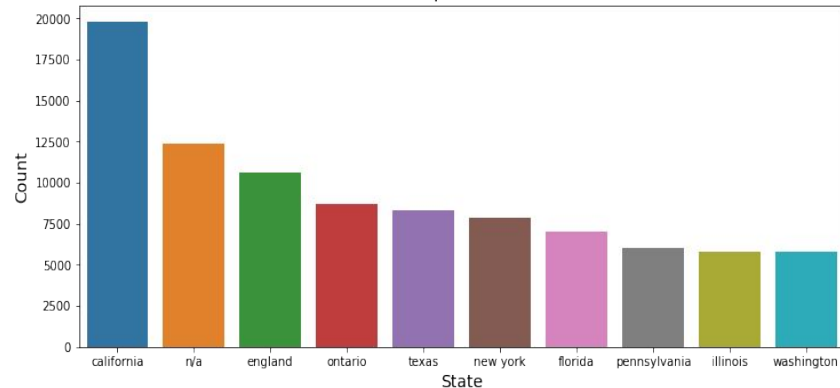


Exploratory Data Analysis

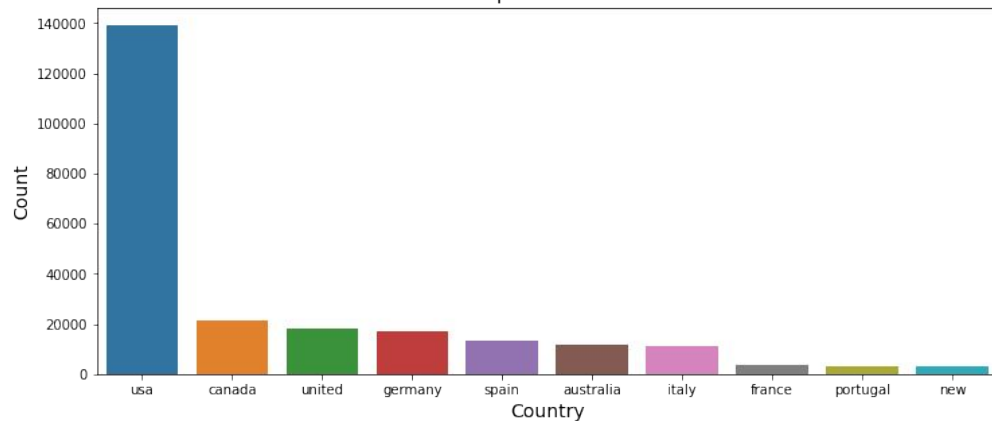
Top Cities



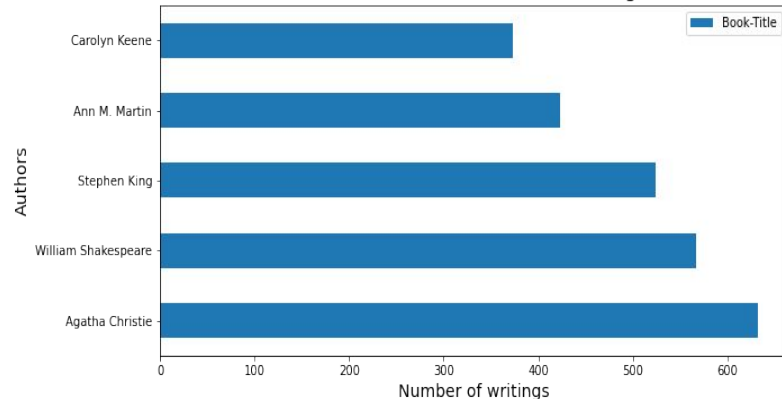
Top States



Top Countries



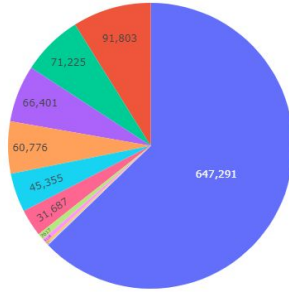
Authors with more number of writings



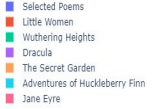
Exploratory Data Analysis



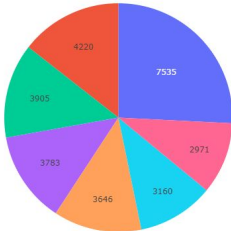
Book Ratings and their percentage of data



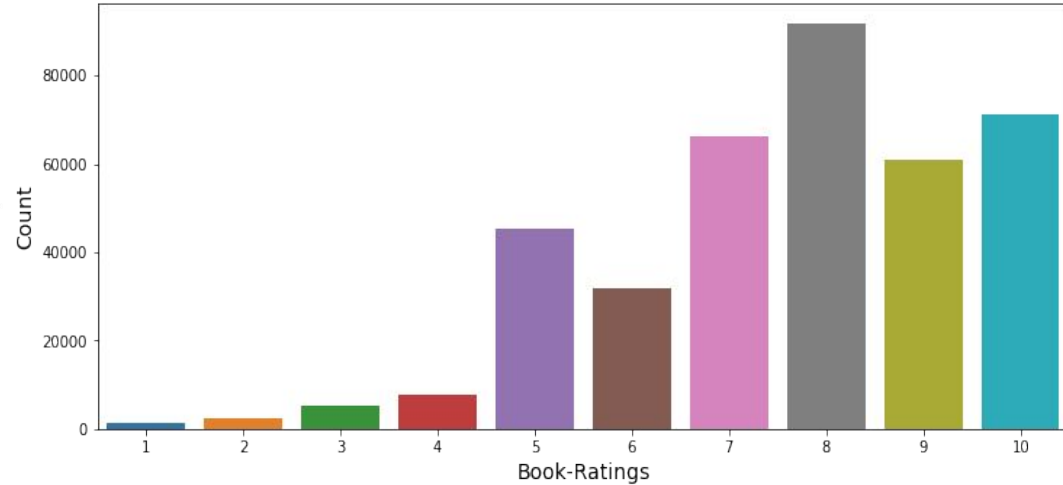
Top Read Books



Top Publishers



Rating given by Users to the Books



Data Preparation & Model Implementation

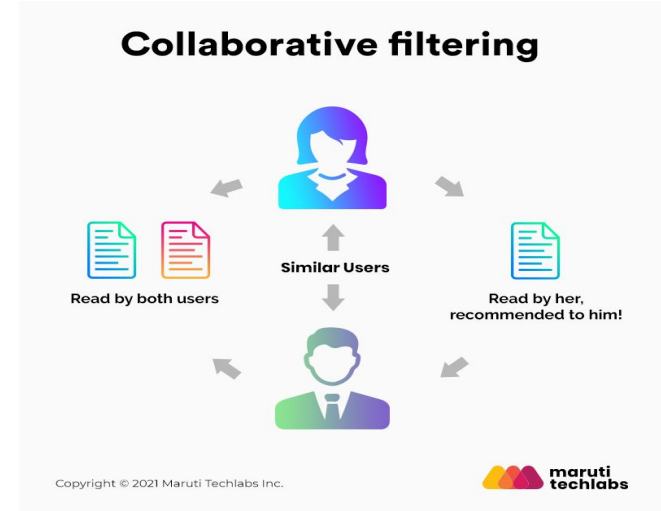
Data Preparation

- Merging datasets.
- Feature selection.
- Calculation of Sparsity.

Recommendation System

Seeks to predict a user's interest in available items and give recommendations accordingly.

- Content Based Filtering
- Collaborative Filtering



	User-ID	ISBN	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher
0	277427	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc
1	11676	002542730X	6	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc
2	12538	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc
3	52584	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc
4	110934	002542730X	6	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc

Popularity Based Recommendation Model



Following are the top 10 Popular Books

	ISBN	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher
0	0316666343	595	The Lovely Bones: A Novel	Alice Sebold	2002	Little, Brown
1	059035342X	411	Harry Potter and the Sorcerer's Stone (Harry P...	J. K. Rowling	1999	Arthur A. Levine Books
2	043935806X	409	Harry Potter and the Order of the Phoenix (Boo...	J. K. Rowling	2003	Scholastic
3	0385504209	400	The Da Vinci Code	Dan Brown	2003	Doubleday
4	0312195516	375	The Red Tent (Bestselling Backlist)	Anita Diamant	1998	Picador USA
5	0446310786	358	To Kill a Mockingbird	Harper Lee	1988	Little Brown & Company
6	0439139597	331	Harry Potter and the Goblet of Fire (Book 4)	J. K. Rowling	2000	Scholastic
7	0345370775	320	Jurassic Park	Michael Crichton	1999	Ballantine Books
8	0439064864	314	Harry Potter and the Chamber of Secrets (Book 2)	J. K. Rowling	1999	Scholastic
9	0439136350	300	Harry Potter and the Prisoner of Azkaban (Book 3)	J. K. Rowling	1999	Scholastic

User Based Recommendation Model

- Rating dataset.
- K Nearest Neighbor model.
- $k = 10$.
- Cosine similarity metric.

```
Predicted rating for user 11676 -> item Legacy: 7  
7
```

```
Predicted rating for user 4017 -> item Life of Pi: 3  
3
```

Item Based Recommendation Model

- Pivot table.
- Index - UserID.
- Columns - Book-Title.
- Values - Book-Rating.

Recommended books similar to Legacy

	Correlation	Ratings-Count
Book-Title		
Harry Potter and the Goblet of Fire (Book 4)	0.096417	54
Harry Potter and the Prisoner of Azkaban (Book 3)	0.076245	58
The Lovely Bones: A Novel	0.040882	71
The Da Vinci Code	0.024978	51
To Kill a Mockingbird	0.022518	52
Harry Potter and the Chamber of Secrets (Book 2)	0.009635	72
Bridget Jones's Diary	0.007517	64

Recommended books similar to Life of Pi

	Correlation	Ratings-Count
Book-Title		
The Lovely Bones: A Novel	0.209224	71
To Kill a Mockingbird	0.136422	52
The Da Vinci Code	0.097563	51
Bridget Jones's Diary	0.061692	64
Harry Potter and the Chamber of Secrets (Book 2)	0.031317	72
Harry Potter and the Prisoner of Azkaban (Book 3)	0.027233	58
Harry Potter and the Goblet of Fire (Book 4)	-0.039437	54

Conclusion

- Most of the users are from the 20 to 40 age group.
- Most of the users are from California, USA.
- Selected poems were read more by users.
- Among authors, Agatha Christie has more writings.
- Harlequin is having more books published.
- 8 was frequently rated by users.
- Popularity based, user-based and item- based recommendation models are created.



Thank you