Book Recommendation System

Roohi Firdous

Content

- Problem Statement
- Data
- Analysis of the datasets
- Data Cleaning
- Treating Outliers
- Imputing Missing Values
- Different Models
- Conclusion

Description

During the last few decades recommendation systems are used widely in amazon, netflix. in terms of providing highly personalized and relevant content

The book recommendation system creates a system for the user based on popularity and user interest.

Dataset Used

01

User dataset

- User ID
- Location
- Age

02

Books Dataset

- ISBN
- Book Title
- Author
- Year of publication
- Publisher

03

Ratings Dataset

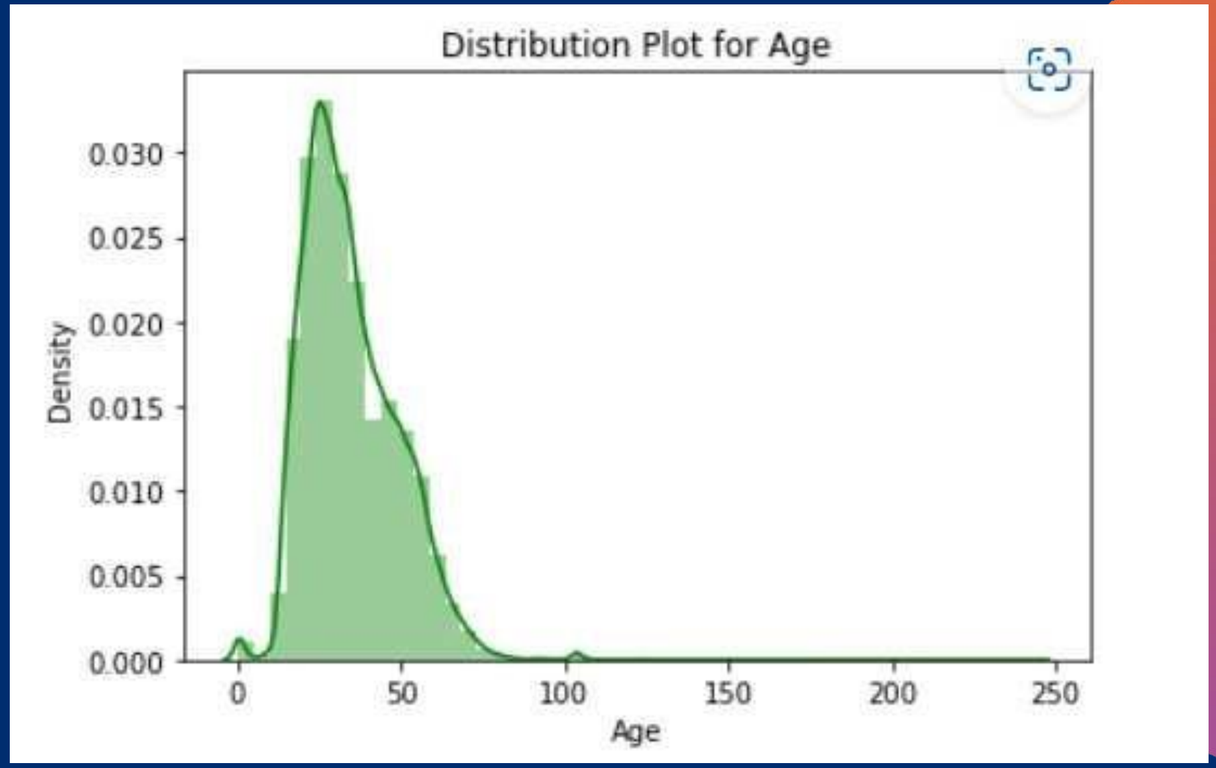
- User ID
- ISBN
- Book Rating

observations From Users df

sns.distplot(users.Age, color='green')

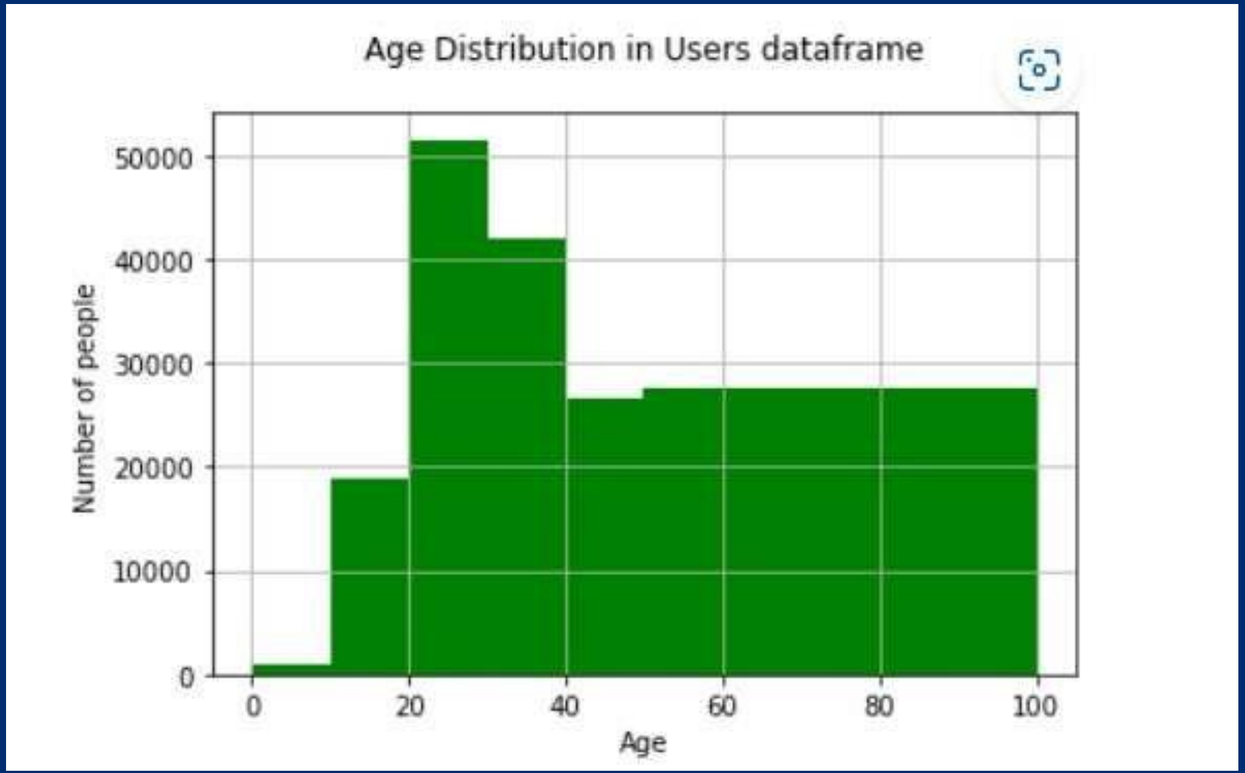
The age range given here is 0-250.

Also the outliers are shown in the plot.
The outliers are right skewed.



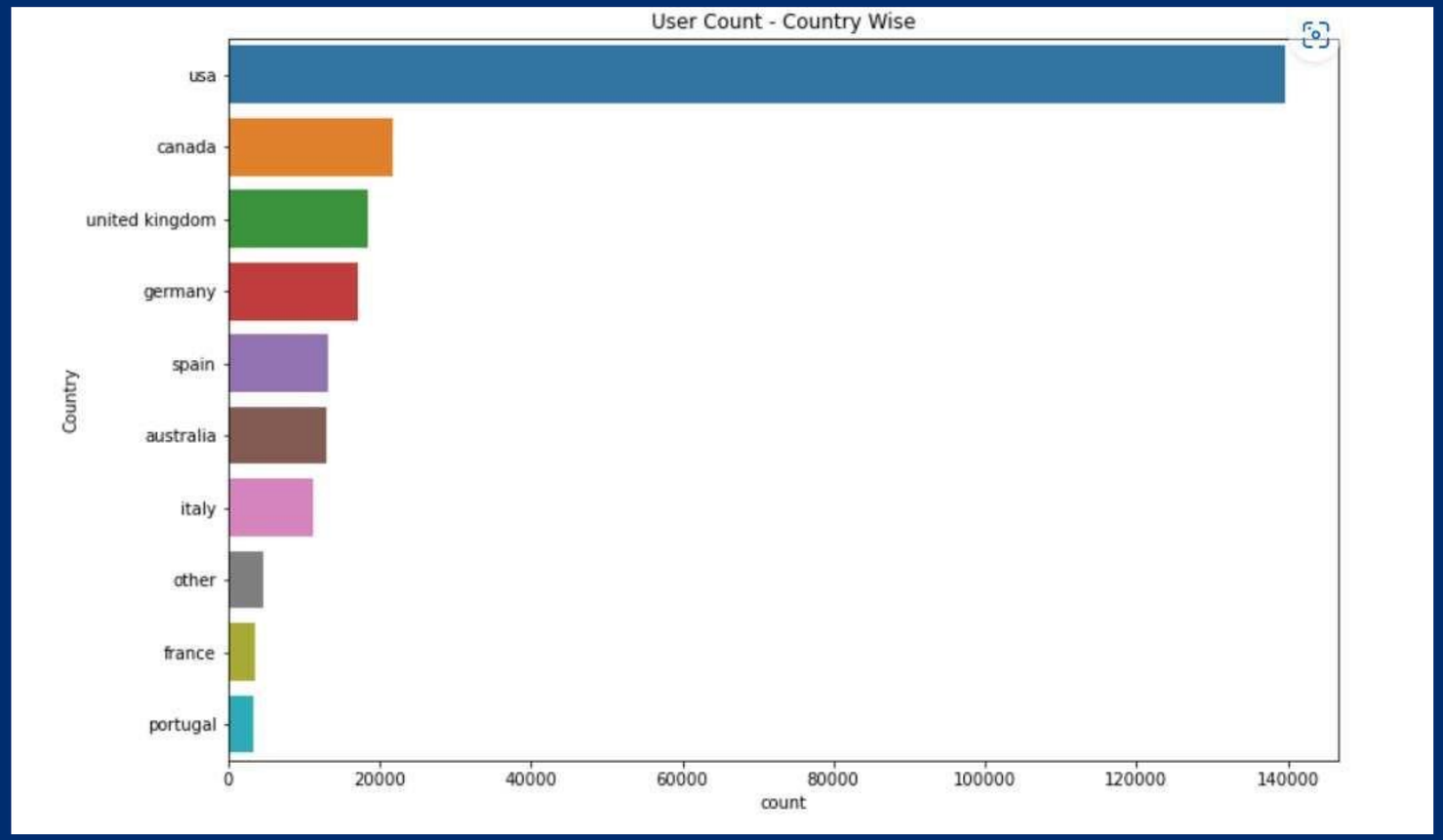
observations From Users df

Most active readers are in the age 20-30



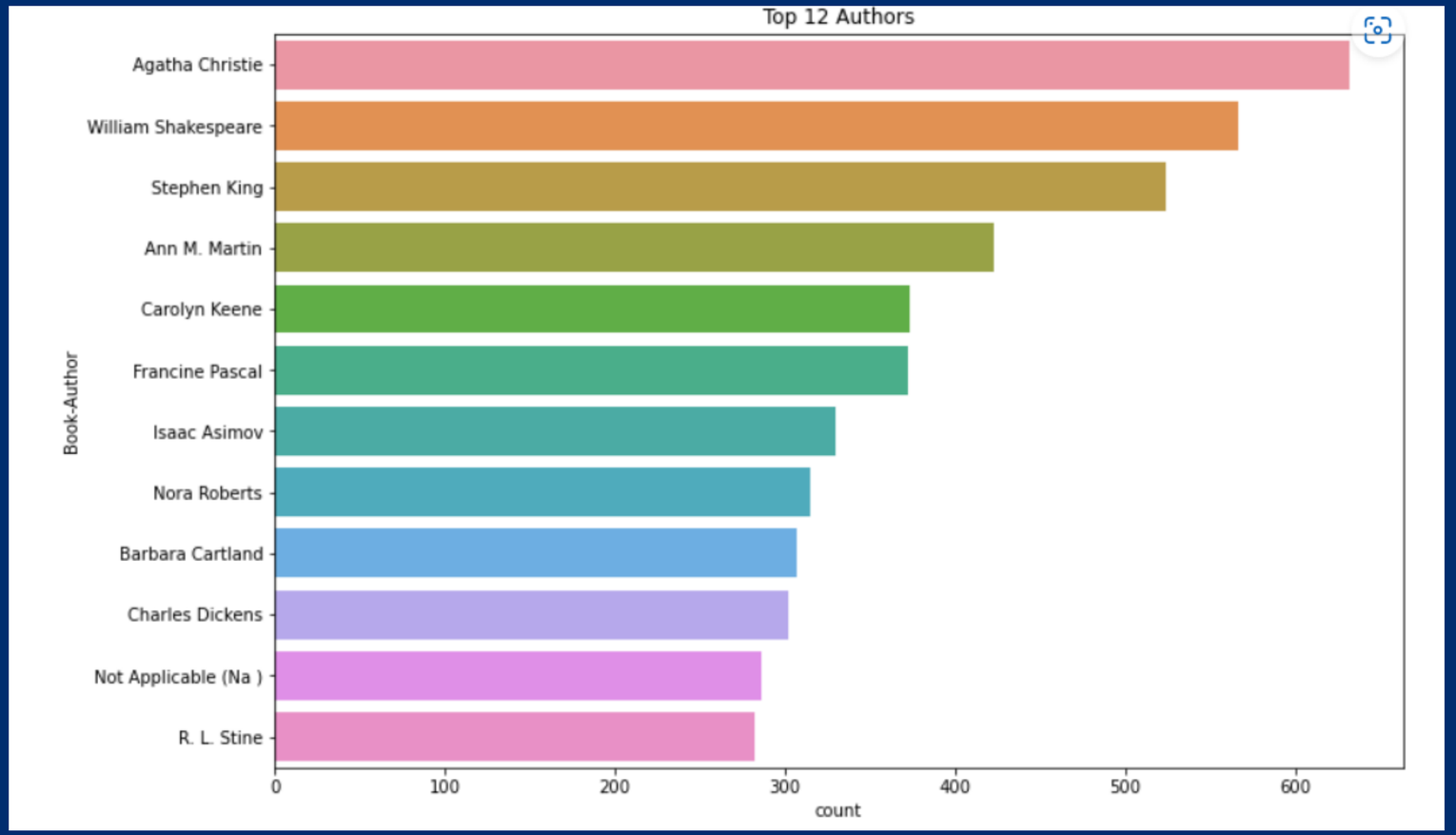
observations From Users df

Splitting the location column and found that the majority of the readers are from Canada and the US



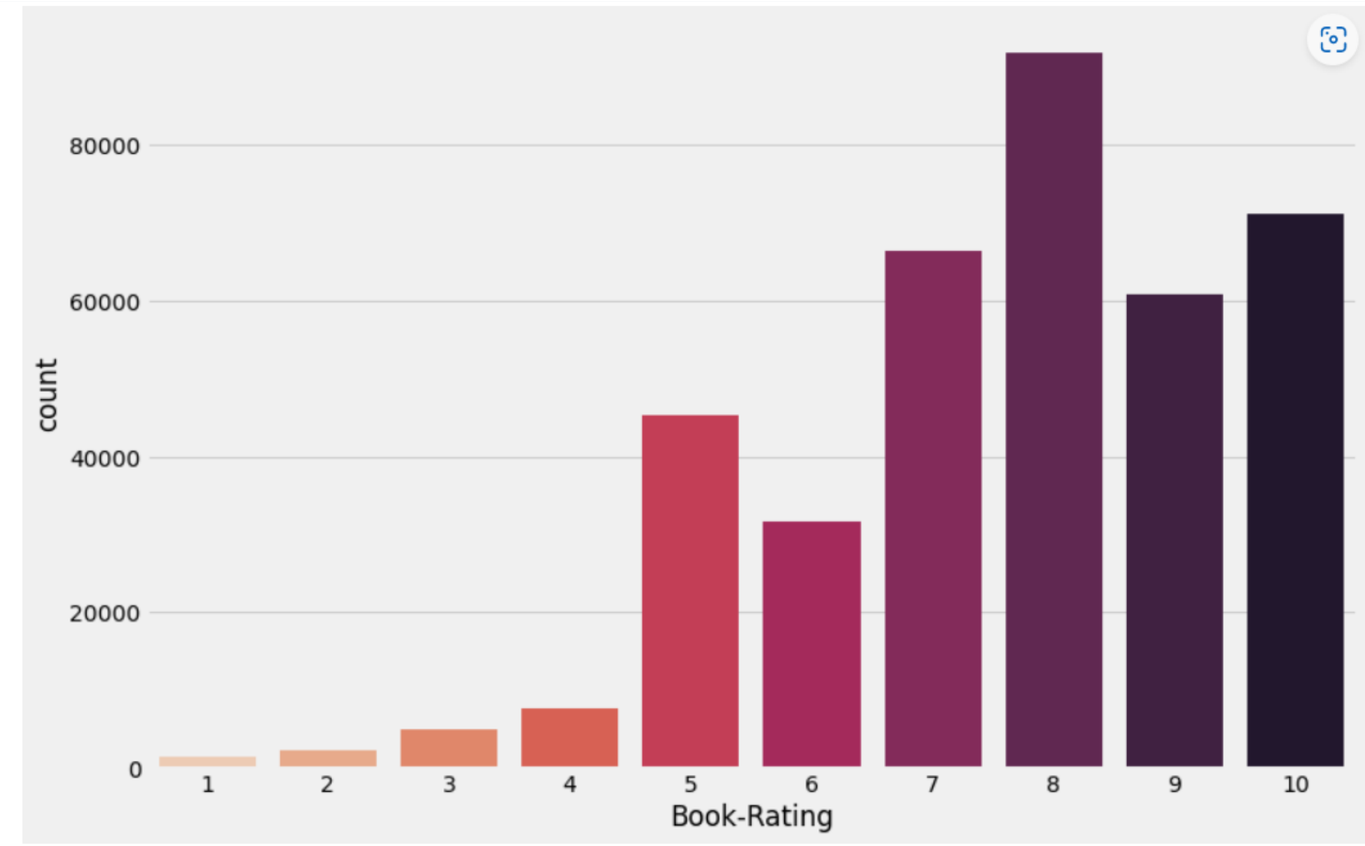
observations From Books df

Agathe Christie wrote the highest number of books as per the dataset.



observations From Ratings df

Higher ratings are common among users.
Rating 8 has been given more number of times.

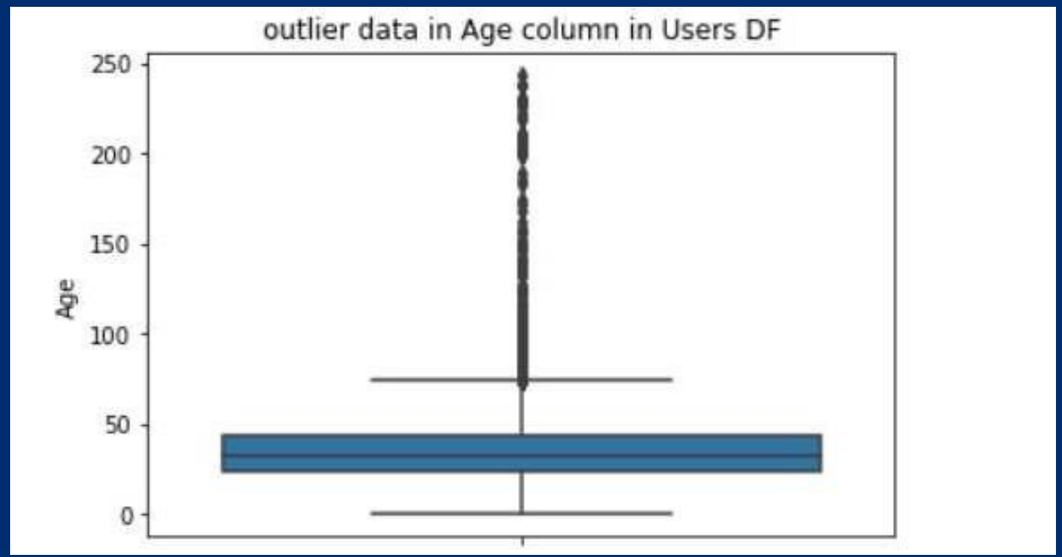


Data Cleaning

1. NULL Value Imputation
   Age column has 40%of NULL values

	index	Missing Values	Percent of Total Values	Data_type
0	Age	110762	39.72	float64
1	User-ID	0	0.00	int64
2	Location	0	0.00	object

2. Outliers in Age column
   We use median to fill NAN values



Models

1. Popularity based Recommendation

Book Weighted Average Formula,

Weighted Rating(WR)= [vR/(v+m)] + [mC/(v+m)]

Where,

v -> the number of votes for the books;
m -> the minimum votes required to be listed in the chart;
R -> the average rating of the book; and
C -> the mean vote across the whole report.

Collaborative Filtering

1. Collaborative Filtering Item-Item Based

   K Nearest Neighbour

```
Recommendations for Blessings : A Novel:

1: Angry Housewives Eating Bon Bons (Ballantine Reader's Circle), with distance of 0.9217835136796061
2: Don't Let's Go to the Dogs Tonight : An African Childhood, with distance of 0.9240185001561373:
3: Peachtree Road, with distance of 0.9389350536368014:
4: The Virgin Blue, with distance of 0.9415693253777184:
5: Creature, with distance of 0.9445250627597287:
```

Collaborative Filtering

1. Collaborative Filtering User-Item Based

```
Enter User ID from above list for book recommendation  11676
Recommendation for User-ID =  11676
         ISBN                          Book-Title  recStrength
0  0385504209                The Da Vinci Code        0.102
1  0452282152         Girl with a Pearl Earring        0.088
2  0312980140  Seven Up (A Stephanie Plum Novel)        0.071
3  0743418174                        Good in Bed        0.068
4  0440212561                          Outlander        0.064
5  0553250531            The Valley of Horses        0.062
6  068484267X          Angela's Ashes: A Memoir        0.061
7  0440214041                The Pelican Brief        0.061
8  0446606812             Message in a Bottle        0.060
9  0440220602                      The Chamber        0.060
```

References

- Datasets downloaded from Kaggle
- https://www.kaggle.com/code/arashnic/recom-i-dataunderstanding-and-simple-recomm/input?select=Books.csv
- https://www.analyticsvidhya.com/blog/2021/06/build-book-recommendation-system-unsupervised-learning-project/

Thank You