CAPSTONE PROJECT-4

Book Recommendation using Collaborative Approach

Unsupervised Learning

Presented By

Rahul Jha

Content:-

- PROBLEM STATEMENT
- APPROACH
- DATA SUMMARY
- DATA VISUALIZATIONS
- FEATURE ENGINEERING
- GENERATING RECOMMENDATIONS
- CONCLUSIONS

Problem Statement

On the Internet, where the number of choices isoverwhelming, there is need to filter, prioritize and efficiently deliver relevant information in order to alleviate the problem of information overload, which has created a potential problem to many Internet users.

Recommender systems solve this problem by searching through large volume of dynamically generated information to provide users with personalized content and services. Therefore, the need to use efficient and accurate recommendation techniques within a system that will provide relevant and dependable recommendations for users cannot be over-emphasized.

Approach

- We examine three datasets in this project: Books.csv, Users.csv, and Ratings.csv. These databases include information about Books, Authors, Publication Year, and Ratings.
- The dataset has been subset due to the enormous volume of data.
- Reduced quantity of records Feature Engineering has been carried out to
- Choose appropriate characteristics and perform a crosstab of the combined data.
- The collaborative recommendation system is then created.
- Filtering (Item Similarity) such that the top ten relevant and
- According to the entered search, strongly connected books are displayed.

Collaborative Approach

- **Collaborative filtering needs a set of items that are based on the user's historical choices. This system does not require a good amount of product features to work. An embedding or feature vector describes each item and User, and it sinks both the items and the users in a similar embedding location. It creates enclosures for items and users on its own. Other purchaser's reactions are taken into consideration while suggesting a specific product to the primary user. It keeps track of the behavior of all users before recommending which item is mostly liked by users. It also relates similar users by similarity in preference and behavior towards a similar product when proposing a product to the primary customer.**

- **Two sources are used to record the interaction of a product user. First, through implicit feedback, User likes and dislikes are recorded and noticed by their actions like clicks, listening to music tracks, searches, purchase records, page views, etc.**

- **Collaborative Filtering is the most famous application suggestion engine and is based on calculated guesses; the people who liked the product will enjoy the same product in the future. This type of algorithm is also known as a product-based collaborative shift. In this Filtering, users are filtered and associated with each User in place of items. In this system, only users' behavior is considered. Only their content and profile information is not enough. The User giving a positive rating to products will be associated with other User's behavior giving a similar rating.**

- **The main idea behind this approach is suggesting new items based on the closeness in the behavior of similar customers.**

Data Overview:-

Books dataset:-

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...

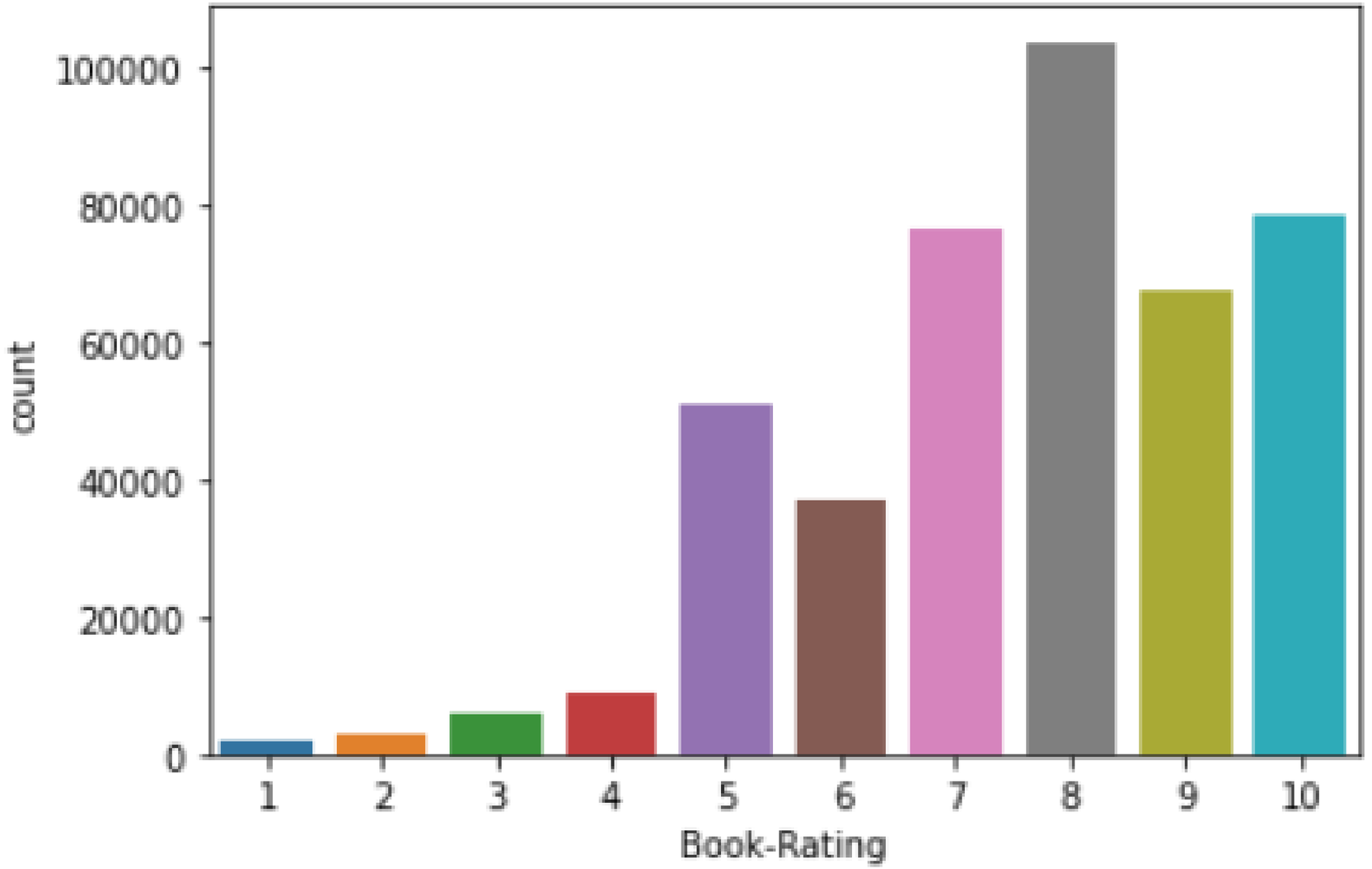
Final Workable Dataset:-

	User-ID	ISBN	Book-Rating_x	Book-Title	Book-Author	Year-Of-Publication	Publisher	Total-no-of-Rating
0	277427	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	82
1	3363	002542730X	0	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	82
2	11676	002542730X	6	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	82
3	12538	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	82

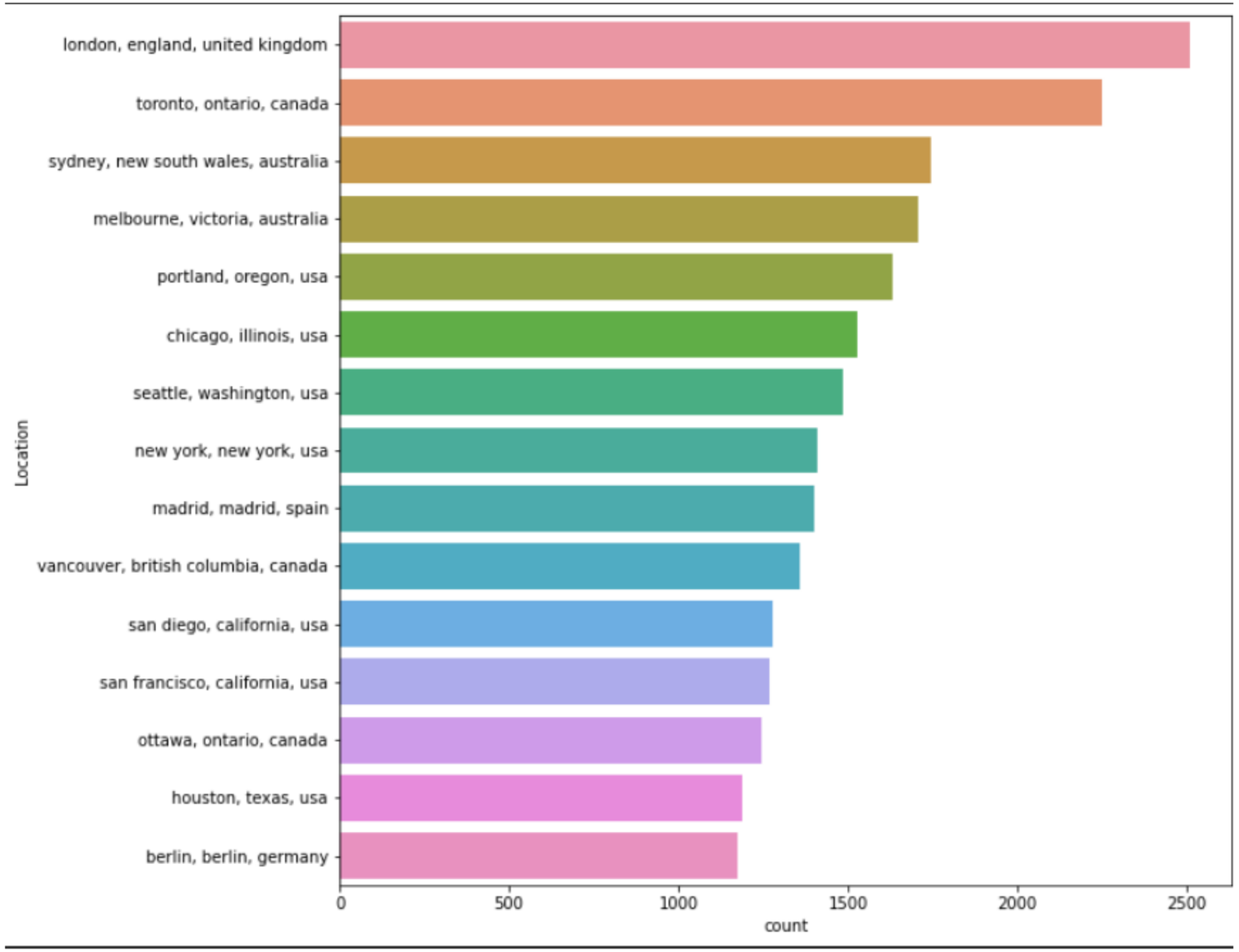
Summary Dataset:-

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher
count	271360	271360	271359	271360	271358
unique	271360	242135	102023	202	16807
top	0195153448	Selected Poems	Agatha Christie	2002	Harlequin
freq	1	27	632	13903	7535

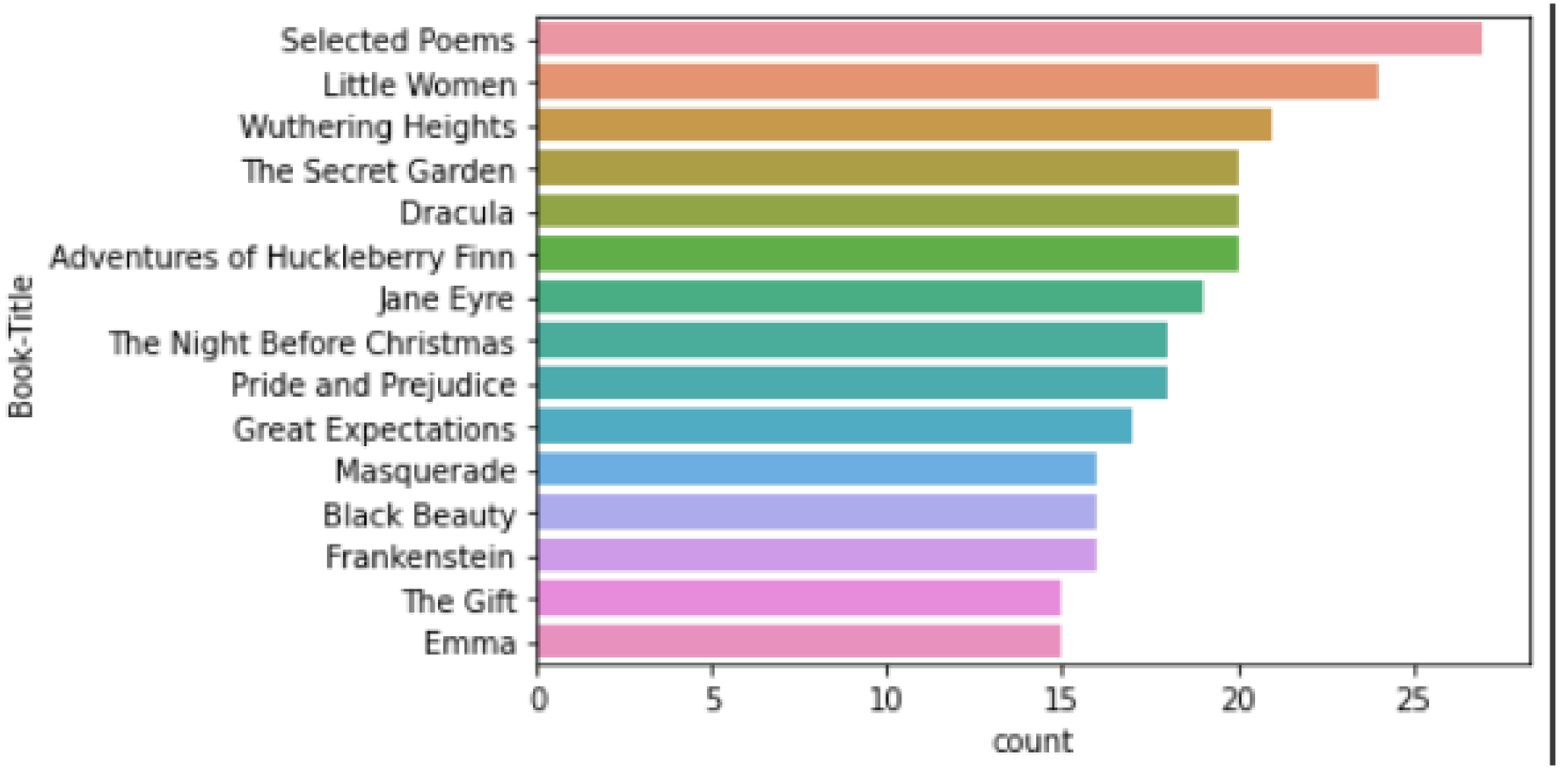
Countplot of book rationg:-



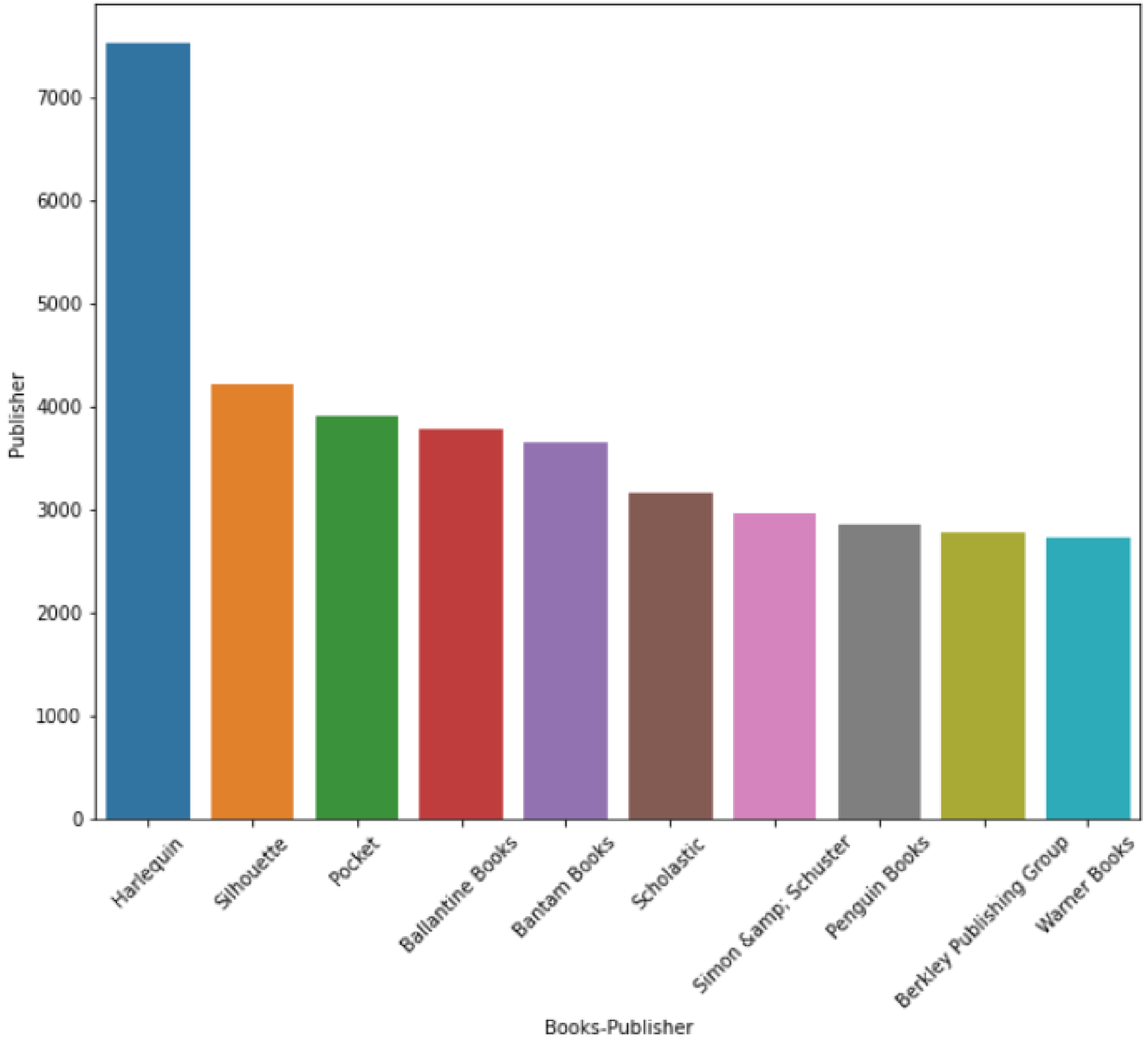
Top location where most user live:-



Top 15 books with most review:-

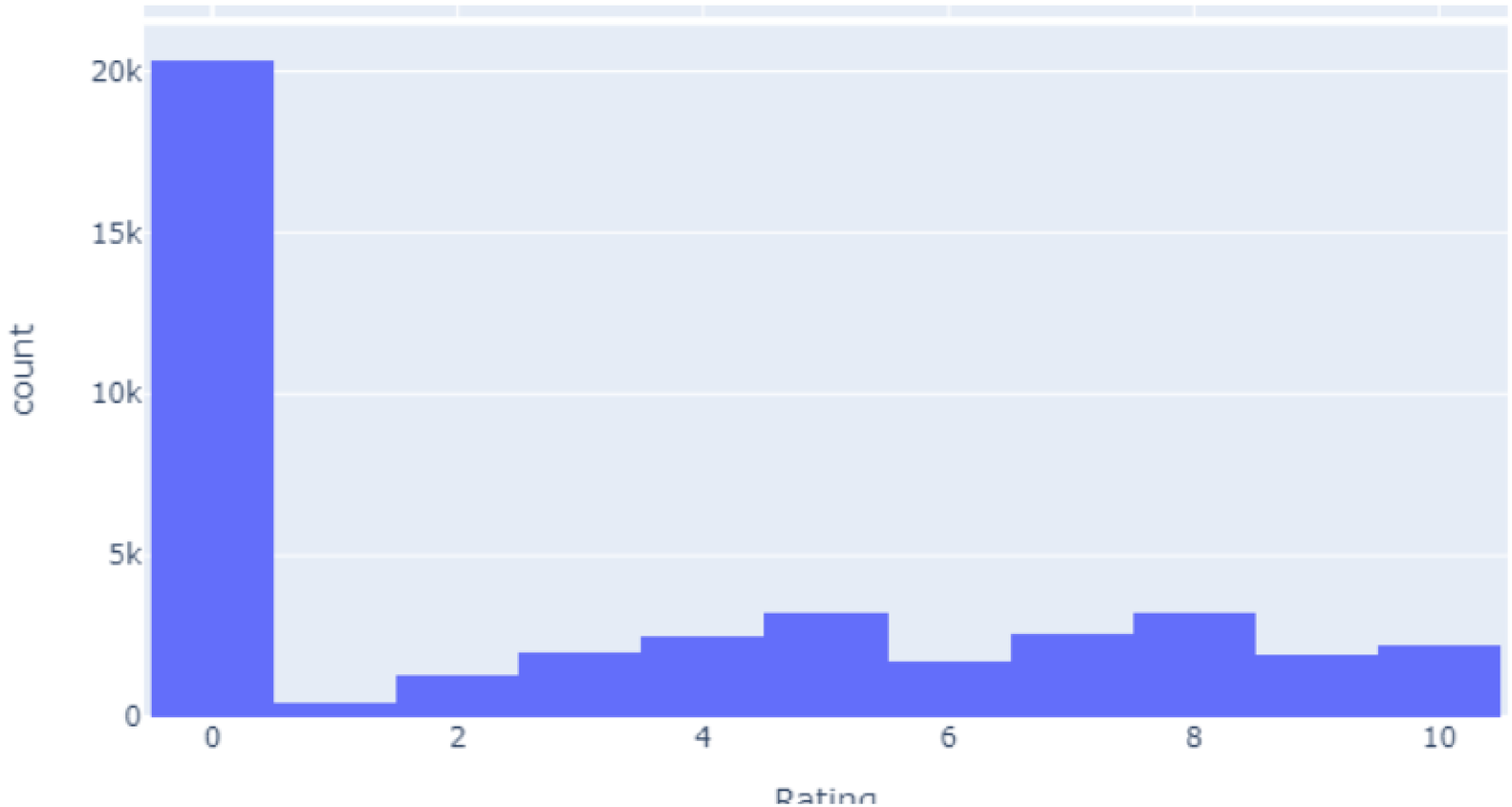


Top 10 book publisher:-



Title	Rating	Num_of_ratings
Wild Animus	0.954315	197
The Lovely Bones: A Novel	4.504673	107
The Secret Life of Bees	5.294872	78
The Da Vinci Code	5.078947	76
Life of Pi	3.985915	71
The Nanny Diaries: A Novel	4.214286	70
Divine Secrets of the Ya-Ya Sisterhood: A Novel	4.000000	70
The Red Tent (Bestselling Backlist)	4.484375	64
A Painted House	3.468750	64
Bridget Jones's Diary	3.269841	63

Average Rate Counting



Rating

Feature Engineering

```
# Create a pivot table/crosstab with the DataFrame
data_pivot = data.pivot_table(index="user_id", columns="Title", values="Rating")
data_pivot.head()
```

Title	A Light in the Storm: The Civil War Diary of Amelia Martin, Fenwick Island, Delaware, 1861 (Dear America)	Beyond IBM: Leadership Marketing and Finance for the 1990s	It Takes Two	Little Comic Shop of Horrors (Give Yourself Goosebumps, Book 17)	Murder of a Sleeping Beauty (Scumble River Mysteries (Paperback))	Q-Space (Star Trek The Next Generation, Book 47)	Tales of Terror and Suspense	$oft Money: The True Power in Our Nation's Capital	' Sie belieben wohl zu scherzen, Mr. Feynman.'. Abenteuer eines neugierigen Physikers.	' Small g'. Eine Sommeridylle.
..										

Book Recommendation

```
def recommend_book(book):
  index=np.where(final.index==book)[0][0]
  distances,suggestion=model.kneighbors(final.iloc[index,:].values.reshape(1,-1),n_neighbors=6)
  for i in range(len(suggestion)):
    print(final.index[suggestion[i]][1:])
```

Book recommendation:- ----------

```
recommend_book('1984')
```

```
Index(['No Safe Place', 'A Civil Action', 'Foucault's Pendulum',
       'Long After Midnight', 'Master of the Game'],
      dtype='object', name='Book-Title')
```

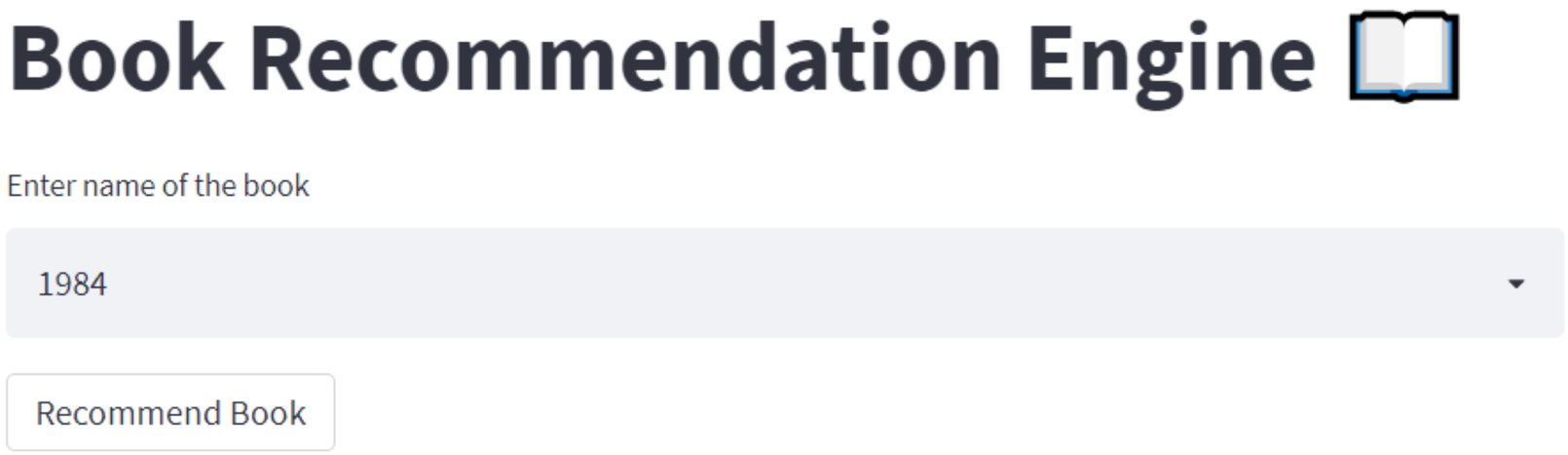
Frontend using Streamlit:-

Streamlit is an open-source python framework for building web apps for Machine Learning and Data Science. We can instantly develop web apps and deploy them easily using Streamlit. Streamlit allows you to write an app the same way you write a python code. Streamlit makes it seamless to work on the interactive loop of coding and viewing results in the web app.

Deployment using Heroku

Heroku’s flexibility allows you to build apps using the language or framework that you know best or love to use the most. It’s a great place to practice popular architectural patterns, such as microservices, or deployment methodologies, such as continuous delivery. When you run into issues, our robust documentation, support, and community resources are available to help you understand the root cause and troubleshoot quickly.

Screenshot of frontend:-



Screenshot:-

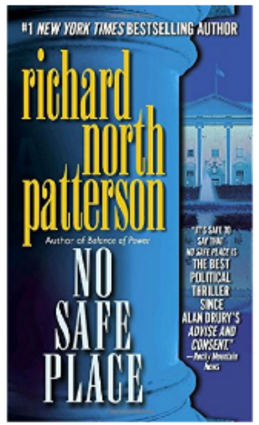
Book Recommendation Engine 📖

Enter name of the book

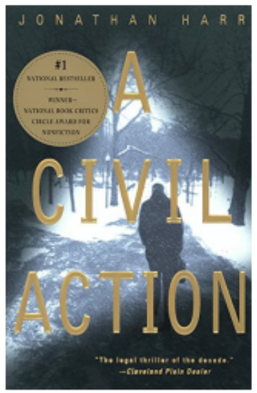
1984

Recommend Book

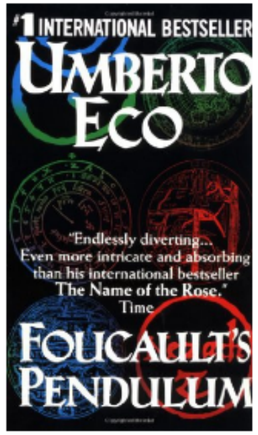
No Safe Place



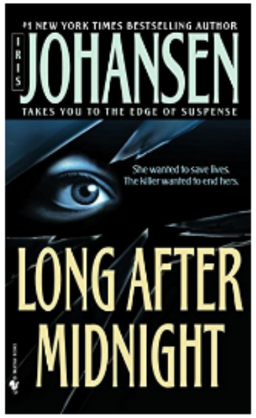
A Civil Action



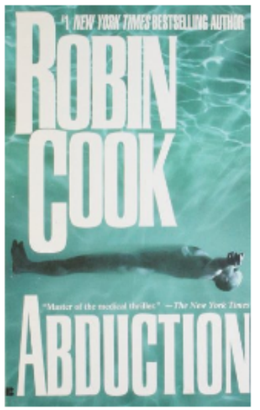
Foucault's Pendulum



Long After Midnight



Abduction



Conclusion:-

- **The top 3 most prolific authors are Stephen King, Nora Roberts and John Grisham.**
- **The most popular novels/books are The Secret Life of Bees, The Da Vinci Code and The Lovely Bones.**
- **The Year in which highest number of book have been published is 2002 and the Publisher with highest number of publications is "Ballantine Books".**
- **The Recommendation system has been implemented using Collaborative filtering and correlation property on unlabeled data. The top 10 most relevant and correlated books can be recommended for any given book in the data using this function. Clustering doesn't help much in this project and hence has not been used.**

THANK YOU: