

Capstone Project

Book Recommendation System

G. V. Kapeesh Varma

CONTENTS

Sr. No.	Topic
1)	PROBLEM STATEMENT
2)	APPROACH
3)	DATA SUMMARY
4)	DATA VISUALIZATIONS
5)	FEATURE ENGINEERING
6)	GENERATING RECOMMENDATIONS
7)	CONCLUSIONS

PROBLEM STATEMENT

On the Internet, where the number of choices is overwhelming, 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

In this project, we analyze three datasets namely Books.csv, Users.csv and Ratings.csv. These datasets altogether provide information about Books, Authors, Year of publication and Ratings. Due to the large amount of data, the dataset has been subset to a lower number of records. Feature Engineering has been performed to select relevant features and crosstab of the merged data. Recommendation system is then designed in such a way that the top 10 relevant and highly correlated books are shown according to the input search.

DATA SUMMARY

	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...

First 5 rows of the Books DataFrame

	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

First 5 rows of the Users DataFrame

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   User-ID     100000 non-null  int64
1   Location    100000 non-null  object
2   Age         60269 non-null   float64
dtypes: float64(1), int64(1), object(1)
memory usage: 2.3+ MB
```

Concise summary of the Users DataFrame

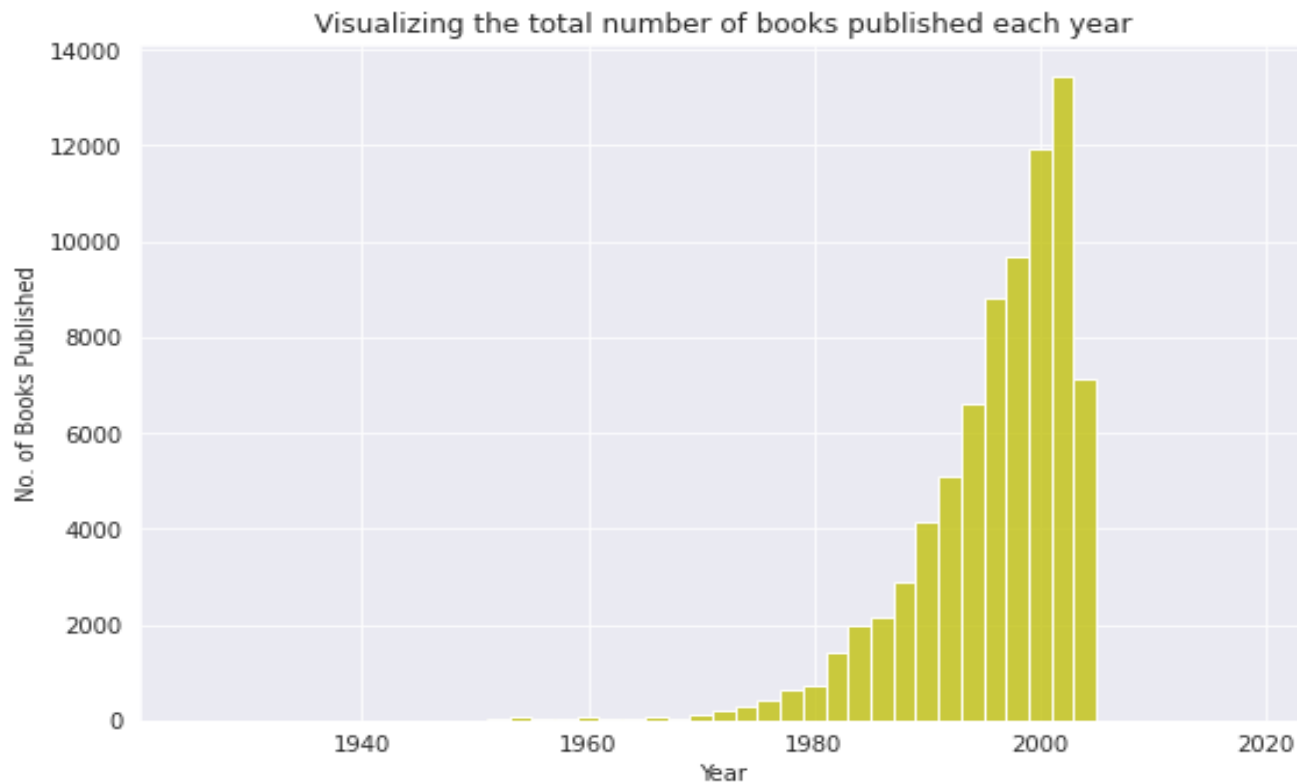
	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

First 5 rows of the Ratings DataFrame

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   User-ID     100000 non-null  int64
1   ISBN        100000 non-null  object
2   Book-Rating 100000 non-null  int64
dtypes: int64(2), object(1)
memory usage: 2.3+ MB
```

Concise summary of the Ratings DataFrame

DATA VISUALIZATIONS



Ballantine Books	2799
Pocket	2455
Berkley Publishing Group	2214
Warner Books	1969
Bantam Books	1828
...	
McGraw Hill (Tx)	1
Harlequin Sales Corp (Mm)	1
B&Hme und Erb	1
FC&A Publishing	1
Publica&Ampes Dom Quixote	1

Name: Publisher, Length: 4434, dtype: int64

**Publishers ranked according to the
total number of publications**

2002	7152
2001	6283
1999	6072
2000	5841
2003	5358
...	
1925	1
1945	1
1927	1
1939	1
1938	1

Name: Year-Of-Publication,

**Years ranked according to the
total number of publications**

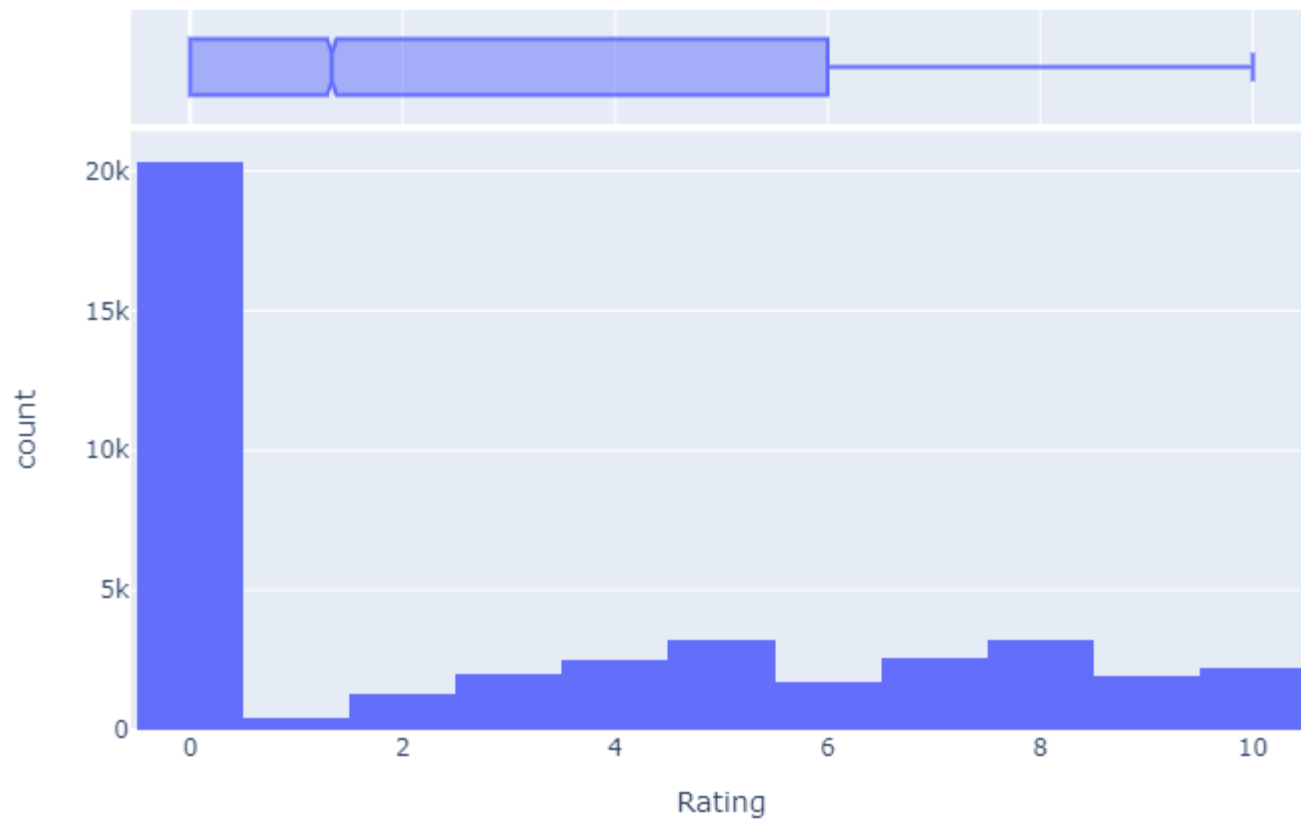
	Rating	Num_of_ratings
Title		
Wild Animus	0.954315	197
The Lovely Bones: A Novel	4.504673	107
The Secret Life of Bees	5.294872	78
The Da Vinci Code	5.078947	76
Life of Pi	3.985915	71
The Nanny Diaries: A Novel	4.214286	70
Divine Secrets of the Ya-Ya Sisterhood: A Novel	4.000000	70
The Red Tent (Bestselling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

Popularity of Books

Author	
Stephen King	741
Nora Roberts	508
John Grisham	486
James Patterson	426
Mary Higgins Clark	404
Dean R. Koontz	357
Tom Clancy	309
Danielle Steel	289
Janet Evanovich	285
Sue Grafton	259

Name: Title, dtype: int64

Top 10 most Prolific Authors



Average Rating count

[illegible]

GENERATING RECOMMENDATIONS



	Title	Correlation	Num_of_ratings
1306	The Da Vinci Code	1.000000	76
1261	The Brethren	1.000000	47
1606	The Summons	0.845626	42
1286	The Client	0.739153	51
1519	The Pelican Brief	0.702524	60
1498	The Nanny Diaries: A Novel	0.633792	70
605	Harry Potter and the Sorcerer's Stone (Harry P...	0.575758	53
1692	Timeline	0.519361	44
640	House of Sand and Fog	0.505445	53
1619	The Testament	0.495832	52

Recommendations for “The Da Vinci Code”

	Title	Correlation	Num_of_ratings
731	Life of Pi	1.000000	71
1225	The Brethren	1.000000	47
1218	The Bonesetter's Daughter	1.000000	42
571	Harry Potter and the Sorcerer's Stone (Harry P...	0.987878	53
1059	She's Come Undone (Oprah's Book Club (Paperback))	0.925641	42
1060	She's Come Undone (Oprah's Book Club)	0.800641	42
1256	The Client	0.726184	51
1149	Summer Sisters	0.705996	52
567	Harry Potter and the Chamber of Secrets (Book 2)	0.583695	48
1501	The Secret Life of Bees	0.559739	78

Recommendations for "Life of Pi"

	Title	Correlation	Num_of_ratings
1904	The Secret Life of Bees	1.000000	78
1935	The Summons	0.906217	42
1326	She's Come Undone (Oprah's Book Club)	0.607815	42
910	Life of Pi	0.559739	71
2030	Timeline	0.461361	44
2155	White Oleander : A Novel	0.444252	53
1643	The Five People You Meet in Heaven	0.371427	41
713	Harry Potter and the Sorcerer's Stone (Harry P...	0.349561	53
1848	The Pilot's Wife : A Novel	0.347149	48
1725	The Joy Luck Club	0.303152	50

Recommendations for “The Secret Life of Bees”

CONCLUSIONS

1. The top 3 most prolific authors are Stephen King, Nora Roberts and John Grisham.
2. The most popular novels/books are The Secret Life of Bees, The Da Vinci Code and The Lovely Bones.
3. The Year in which highest number of book have been published is 2002 and the Publisher with highest number of publications is "Ballantine Books".
4. The Recommendation system has been implemented using 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.