

Recommendation system for Amazon book store

```
In [203]: import pandas as pd
import numpy as np
import os, sys

import warnings
warnings.filterwarnings('ignore')
```

```
In [200]: Books = pd.read_csv('BX_Books.csv', encoding='latin-1')
ratings = pd.read_csv('BX-Book-Ratings.csv', encoding='latin-1')
users = pd.read_csv('BX-Users.csv', encoding='latin-1')
```

```
In [205]: df
```

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	Image-URL-L	Unnamed: 8	Unnamed: 9	...	Unnamed: 12	Unnamed: 13	Unnamed: 14	Unman
0	195153448	Classical Mythology	Mark P. O. Morford	2002.0	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...	http://images.amazon.com/images/P/0195153448.0...	http://images.amazon.com/images/P/0195153448.0...	NaN	NaN	...	NaN	NaN	NaN	
1	20050518	Clara Callan	Richard Bruce Wright	2001.0	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...	http://images.amazon.com/images/P/0002005018.0...	http://images.amazon.com/images/P/0002005018.0...	NaN	NaN	...	NaN	NaN	NaN	
2	60873129	Decision in Normandy	Carlo D'Este	1991.0	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...	http://images.amazon.com/images/P/0060973129.0...	http://images.amazon.com/images/P/0060973129.0...	NaN	NaN	...	NaN	NaN	NaN	
3	374157065	Fu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999.0	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...	http://images.amazon.com/images/P/0374157065.0...	http://images.amazon.com/images/P/0374157065.0...	NaN	NaN	...	NaN	NaN	NaN	
4	393045218	The Mummies of Urukchi	E. J. W. Barber	1999.0	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...	http://images.amazon.com/images/P/0393045218.0...	http://images.amazon.com/images/P/0393045218.0...	NaN	NaN	...	NaN	NaN	NaN	
...
271374	44400988	There's a Bat in Bunk Five	Paula Danziger	1988.0	Random House Childrens Pub (Mn)	http://images.amazon.com/images/P/0440400988.0...	http://images.amazon.com/images/P/0440400988.0...	http://images.amazon.com/images/P/0440400988.0...	NaN	NaN	...	NaN	NaN	NaN	
271375	525447644	From One Hundred	Teri Sloat	1991.0	Dutton Books	http://images.amazon.com/images/P/0525447644.0...	http://images.amazon.com/images/P/0525447644.0...	http://images.amazon.com/images/P/0525447644.0...	NaN	NaN	...	NaN	NaN	NaN	
271376	006008667X	Lily Dale: The True Story of the Town that	Christine Wicker	2004.0	HarperSanFrancisco	http://images.amazon.com/images/P/006008667X.0...	http://images.amazon.com/images/P/006008667X.0...	http://images.amazon.com/images/P/006008667X.0...	NaN	NaN	...	NaN	NaN	NaN	
271377	192126040	Republic (World's Classics)	Plato	1996.0	Oxford University Press	http://images.amazon.com/images/P/0192126040.0...	http://images.amazon.com/images/P/0192126040.0...	http://images.amazon.com/images/P/0192126040.0...	NaN	NaN	...	NaN	NaN	NaN	
271378	767409752	A Guided Tour of Rene Descartes Meditations o...	Christopher Biffe	2000.0	McGraw-Hill Humanities/Social Sciences/Languages	http://images.amazon.com/images/P/0767409752.0...	http://images.amazon.com/images/P/0767409752.0...	http://images.amazon.com/images/P/0767409752.0...	NaN	NaN	...	NaN	NaN	NaN	

271379 rows × 22 columns

```
In [8]: Books.head(2)
ratings.head(2)
users.head(2)
```

	User-ID	Location	Age	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8
0	1.0	nyc	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2.0	stockton	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
In [9]: print(Books.shape)
print('*****')
print(ratings.shape)
print('*****')
print(users.shape)
```

(271379, 22)

(1048575, 3)

(276271, 9)

```
In [16]: # removing unwanted columns
Books = Books.iloc[:,0:8]
users = users.iloc[:,0:3]
```

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0.0
1	276726	150601224	5.0
2	276727	446520902	0.0
3	276729	052169515X	3.0
4	276729	521795028	6.0
...
1048570	250764	451410777	0.0
1048571	250764	452264464	8.0
1048572	250764	048623715X	0.0
1048573	250764	486256588	0.0
1048574	250764	515069434	0.0

1048575 rows × 3 columns

```
In [17]: print(Books.isnull().sum())
print('*****')
print(ratings.isnull().sum())
print('*****')
print(users.isnull().sum())
```

ISBN 0
Book-Title 0
Book-Author 37632
Year-Of-Publication 38973
Publisher 38975
Image-URL-S 43235
Image-URL-M 43235
Image-URL-L 43235
dtype: int64

User-ID 0
ISBN 1
Book-Rating 5
dtype: int64

User-ID 1
Location 553
Age 276270
dtype: int64

```
In [23]: #percent of null values in book table
print((Books.isnull().sum()/Books.count())*100)
```

ISBN 0.000000
Book-Title 0.000000
Book-Author 16.099458
Year-Of-Publication 16.769361
Publisher 16.770365
Image-URL-S 18.950759
Image-URL-M 18.950759
Image-URL-L 18.950759
dtype: float64

```
In [26]: print(Books.duplicated().sum())
print('*****')
print(ratings.duplicated().sum())
print('*****')
print(users.duplicated().sum())
```

0

0

0

```
In [25]: #dropping duplicates
ratings.drop_duplicates(inplace=True)
```

```
In [ ]: #popular based recommendation system
```

```
In [31]: ratings_with_name = ratings.merge(Books, on='ISBN')
ratings_with_name
```

	User-ID	ISBN	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	Image-URL-L
0	276725	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
1	2313	034545104X	5.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
2	6543	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
3	8690	034545104X	5.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
4	10314	034545104X	9.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
...
941136	250764	440106575	0.0	HIDDEN FIRES	JANETTE RADCLIFFE	1978.0	Dell	http://images.amazon.com/images/P/0440106575.0...	http://images.amazon.com/images/P/0440106575.0...	http://images.amazon.com/images/P/0440106575.0...
941137	250764	451157516	0.0	Cheyenne (Fortunes West	NaN	NaN	NaN	NaN	NaN	NaN
941138	250764	048623715X	0.0	Glamorous Movie Stars of the Thirties: Paper Dolls in Pu...	Tom Tierney	1982.0	Dover Publications	http://images.amazon.com/images/P/048623715X.0...	http://images.amazon.com/images/P/048623715X.0...	http://images.amazon.com/images/P/048623715X.0...
941139	250764	486256588	0.0	Schiaparelli Fashion Review: Paper Dolls in Pa...	Tom Tierney	1988.0	Dover Publications	http://images.amazon.com/images/P/0486256588.0...	http://images.amazon.com/images/P/0486256588.0...	http://images.amazon.com/images/P/0486256588.0...
941140	250764	515069434	0.0	Lady Laughing Eyes (To Have and to Hold)	Lee Damon	1964.0	Jove Books	http://images.amazon.com/images/P/0515069434.0...	http://images.amazon.com/images/P/0515069434.0...	http://images.amazon.com/images/P/0515069434.0...

941141 rows × 10 columns

```
In [37]: num_ratings = ratings_with_name.groupby('Book-Title').count()['Book-Rating'].reset_index()
```

```
In [40]: num_ratings.rename(columns={'Book-Rating':'Num_ratings'}, inplace=True)
num_ratings
```

	Book-Title	Num_ratings
0	A Light in the Storm: The Civil War Diary of ...	4
1	Always Have Popsicles	1
2	Apple Magic (The Collector's series)	1
3	Beyond IBM: Leadership Marketing and Finance ...	1
4	Clifford Vista El Hospital (Clifford El Gran...	1
...
226804	ÃĀÃĀn- Connection.	1
226805	ÃĀÃĀnpiraten.	2
226806	ÃĀÃĀnger mit Produkt X. Roman.	4
226807	ÃĀÃĀnlich der Berge.	2
226808	ÃĀÃĀnrique en toc	2

226809 rows × 2 columns

```
In [46]: avg_ratings = ratings_with_name.groupby('Book-Title').mean()['Book-Rating'].reset_index()
avg_ratings.rename(columns={'Book-Rating':'Avg_ratings'}, inplace=True)
```

```
In [57]: popular_df = num_ratings.merge(avg_ratings, on='Book-Title')
popular_df
```

	Book-Title	Num_ratings	Avg_ratings
0	A Light in the Storm: The Civil War Diary of ...	4	2.25
1	Always Have Popsicles	1	0.00
2	Apple Magic (The Collector's series)	1	0.00
3	Beyond IBM: Leadership Marketing and Finance ...	1	0.00
4	Clifford Vista El Hospital (Clifford El Gran...	1	0.00
...
226804	ÃĀÃĀn- Connection.	1	0.00
226805	ÃĀÃĀnpiraten.	2	0.00
226806	ÃĀÃĀnger mit Produkt X. Roman.	4	5.25
226807	ÃĀÃĀnlich der Berge.	2	4.00
226808	ÃĀÃĀnrique en toc	2	4.00

226809 rows × 3 columns

```
In [58]: popular_df = popular_df.merge(Books, on='Book-Title').drop_duplicates()[['Book-Title', 'Book-Author',
'Year-Of-Publication', 'Publisher',
'Image-URL-S', 'Num_ratings', 'Avg_ratings']]
```

```
In [59]: popular_df
```

	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Num_ratings	Avg_ratings
0	A Light in the Storm: The Civil War Diary of ...	NaN	NaN	NaN	NaN	4	2.25
1	Always Have Popsicles	Rebecca Harvin	1994.0	Rebecca L. Harvin	http://images.amazon.com/images/P/0964147726.0...	1	0.00
2	Apple Magic (The Collector's series)	Marina Boudreau	1984.0	Amner Cooking Guild	http://images.amazon.com/images/P/0942320093.0...	1	0.00
3	Beyond IBM: Leadership Marketing and Finance ...	Lou Mobley	1989.0	Teleonet	NaN	1	0.00
4	Clifford Vista El Hospital (Clifford El Gran...	Norman Bridwell	2000.0	Scholastic	http://images.amazon.com/images/P/0439188970.0...	1	0.00
...
259653	ÃĀÃĀn- Connection.	Heinz GÃĀÃĀnther Korsalk	1995.0	Goldmann	http://images.amazon.com/images/P/3442429617.0...	1	0.00
259654	ÃĀÃĀnpiraten.	Janwillem van de Veleering	2002.0	Rowohlt	http://images.amazon.com/images/P/3499232499.0...	2	0.00
259655	ÃĀÃĀnger mit Produkt X. Roman.	Joan Aiken	1987.0	Diogenes Verlag	http://images.amazon.com/images/P/325721538X.0...	4	5.25
259656	ÃĀÃĀnlich der Berge.	David Guterson	2000.0	btb	http://images.amazon.com/images/P/3442725739.0...	2	4.00
259657	ÃĀÃĀnrique en toc	Didier Darnenckx	1998.0	Balene	http://images.amazon.com/images/P/2842192508.0...	2	4.00

259658 rows × 7 columns

```
In [ ]: #collabrative recommendation sysytem
```

```
In [82]: x = ratings_with_name.groupby('User-ID').count()[['Book-Rating']>250]
x
```

	User-ID	Book-Rating
2	False	
8	False	
9	False	
10	False	
12	False	
278846	False	
278849	False	
278951	False	
278852	False	
278854	False	
Name: Book-Rating, Length: 83644, dtype: bool		

```
In [70]: freq_readers = x[x].index
freq_readers
```

	User-ID	Book-Rating
Int64Index	[254, 2276, 2766, 3363, 4385, 6251, 6543, 6575, 7158, 7346, ... 247752, 248718, 249628, 249862, 249894, 250184, 250405, 277427, 277639, 278418],	
	dtype='int64', name='User-ID', length=554)	

```
In [72]: filter_ratings = ratings_with_name[ratings_with_name['User-ID'].isin(freq_readers)]
filter_ratings
```

	User-ID	ISBN	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	Image-URL-L
2	6543	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
5	28328	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
7	26523	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
15	77940	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
16	81977	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002.0	Ballantine Books	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...	http://images.amazon.com/images/P/034545104X.0...
...
940952	250405	1862071713	0.0	Remind Me Who I Am	NaN	NaN	NaN	NaN	NaN	NaN
940953	250405	1879557487	0.0	Evolution & Extinction: Essays	Stephen Jay Gould	1998.0	Audio Scholar	http://images.amazon.com/images/P/1879557487.0...	http://images.amazon.com/images/P/1879557487.0...	http://images.amazon.com/images/P/1879557487.0...
940954	250405	188424405X	8.0	Menopause	NaN	NaN	NaN	NaN	NaN	NaN
940955	250405	1884433847	0.0	Life Launch: A Passionate Guide to the Rest of...	Frederic M. Hudson	2000.0	Hudson Pr	http://images.amazon.com/images/P/1884433847.0...	http://images.amazon.com/images/P/1884433847.0...	http://images.amazon.com/images/P/1884433847.0...
940956	250405	188838719X	10.0	You've Got a Friend	Dan Zadra	1997.0	Compendium Publishing & Communications	http://images.amazon.com/images/P/188838719X.0...	http://images.amazon.com/images/P/188838719X.0...	http://images.amazon.com/images/P/188838719X.0...

393690 rows × 10 columns

```
In [89]: y = filter_ratings.groupby('Book-Title').count()[['Book-Rating']]>= 50
famous_books = y[y].index
```

```
In [89]: famous_books
```

```
Out[89]: Index(['1984', '1st to Die: A Novel', '2nd Chance', '4 Blondes',
'A Bend in the Road', 'A Case of Need',
'A Child Called It*', 'One Child's Courage to Survive',
'A Is for Alibi (Kinsey Millhone Mysteries (Paperback))',
'A Man in Full', 'A Map of the World',
'White Oleander : A Novel (Oprah's Book Club)',
'Wicked: The Life and Times of the Wicked Witch of the West',
'Wild Animals', 'Winter Solstice', 'Wish You Well', 'Without Remorse',
'Wuthering Heights', 'You Belong To Me',
'Zen and the Art of Motorcycle Maintenance: An Inquiry into Values',
'XOY: Is for Outlaw'], name='Book-Title', length=484)
```

```
In [84]: #final_ratings
final_ratings = filter_ratings[filter_ratings['Book-Title'].isin(famous_books)]
```

```
In [92]: pt = final_ratings.pivot_table(index='Book-Title', columns='User-ID', values='Book-Rating')
```

```
In [85]: pt.fillna(0, inplace=True)
```

```
In [115]:
```

```
.....:
IndexError                                Traceback (most recent call last)
Input In [115], in <cell line 3>()
----> 3 sorted(list((similarity_sc[1984])))
IndexError: index 1984 is out of bounds for axis 0 with size 484
```

```
In [97]: #recommendation system
from sklearn.metrics.pairwise import cosine_similarity
```

```
In [98]: similarity_sc = cosine_similarity(pt)
```

```
In [206]: index = np.where(pt.index=="2nd Chance")[0][0]
index
pt.head(5)
```

	User-ID	254	2276	2766	3363	4385	6251	6543	6575	7158	7346	...	247752	248718	249628	249862	249894	250184	250405	277427	277639	278418
1984	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1st to Die: A Novel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2nd Chance	0.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4 Blondes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A Bend in the Road	0.0	0.0	7.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 553 columns

```
In [195]: def recommendation_system(book_name):
    index = np.where(pt.index==book_name)[0][0]
    similar_title = sorted(list(enumerate(similarity_sc[index])),key= lambda x : x[1],reverse=True)[1:5])

    data = []
    for i in similar_title:
        item = []
        temp_df = Books[Books['Book-Title']==pt.index[i][0]]
        item.extend(list(temp_df.drop_duplicates('Book-Title')['Book-Title'].values))
        data.append(item)
    print(index)
    return data
```

```
In [127]: recommendation_system('1984')
```

0

```
Out[127]: [['Brave New World'],
['The Vampire Lestat ( Vampire Chronicles '],
['The Tale of the Body Thief (Vampire Chronicles (Paperback))'],
['The Handmaid's Tale']]
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```