

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [2]: Books = pd.read_csv("Books112.csv")
users=pd.read_csv("Users112.csv")
Ratings=pd.read_csv("Ratings112.csv")
```

```
C:\Users\91760\AppData\Local\Temp\ipykernel_51664\435766521.py:1: DtypeWarning: Columns (3) have mixed types. Specify dtype option on import or set low_memory=False.
  Books = pd.read_csv("Books112.csv")
C:\Users\91760\AppData\Local\Temp\ipykernel_51664\435766521.py:2: DtypeWarning: Columns (0) have mixed types. Specify dtype option on import or set low_memory=False.
  users=pd.read_csv("Users112.csv")
```

```
In [3]: Books.head(5)
```

```
Out[3]:
```

	isbn	book_title	book_author	year_of_publication	publisher
0	195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press
1	2005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada
2	60973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial
3	374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux
4	393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company

```
In [4]: Ratings.head(6)
```

```
Out[4]:
```

	user_id	isbn	rating
0	276725	034545104X	0.0
1	276726	155061224	5.0
2	276727	446520802	0.0
3	276729	052165615X	3.0
4	276729	521795028	6.0
5	276733	2080674722	0.0

```
In [5]: users.head(6)
```

```
Out[5]:
```

	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
5	6	santa monica, california, usa	61.0

```
In [6]: users.shape
```

```
Out[6]: (278859, 3)
```

```
In [7]: print(Ratings.shape)
```

```
(1048575, 3)
```

```
In [8]: print(Books.shape)
```

```
(271379, 5)
```

```
In [9]: Books.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271379 entries, 0 to 271378
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0   isbn                  271379 non-null  object
1   book_title            271379 non-null  object
2   book_author           271378 non-null  object
3   year_of_publication    271379 non-null  object
4   publisher              271377 non-null  object
dtypes: object(5)
memory usage: 10.4+ MB
```

```
In [10]: Ratings.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   user_id     1048575 non-null  int64
1   isbn        1048575 non-null  object
2   rating      1048573 non-null  float64
dtypes: float64(1), int64(1), object(1)
memory usage: 24.0+ MB
```

```
In [11]: users.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 278859 entries, 0 to 278858
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   user_id     278859 non-null  object
1   Location     278858 non-null  object
2   Age          168096 non-null  float64
dtypes: float64(1), object(2)
memory usage: 6.4+ MB
```

```
In [12]: Books[Books.duplicated()]
```

```
Out[12]:
```

isbn	book_title	book_author	year_of_publication	publisher
------	------------	-------------	---------------------	-----------

```
In [13]: users[users.duplicated()]
```

```
Out[13]:
```

user_id	Location	Age
---------	----------	-----

```
In [14]: Ratings[Ratings.duplicated()]
```

```
Out[14]:
```

	user_id	isbn	rating
11338	709	9.78E+12	9.0
21604	4334	6.31E+11	0.0
28622	6575	6.31E+11	0.0
58876	11676	9.78E+12	9.0
58877	11676	9.78E+12	9.0
...

1047765	250634	9.78E+12	0.0
1047766	250634	9.78E+12	10.0
1047767	250634	9.78E+12	10.0
1047768	250634	9.78E+12	10.0
1047769	250634	9.78E+12	0.0

146 rows × 3 columns

```
In [15]: Ratings.drop_duplicates(inplace=True)
Ratings.reset_index(drop=True, inplace=True)
Ratings.shape
```

```
Out[15]: (1048429, 3)
```

```
In [16]: Books.isnull().sum()
```

```
Out[16]: isbn                0
book_title              0
book_author             1
year_of_publication     0
publisher               2
dtype: int64
```

```
In [17]: Ratings.isnull().sum()
```

```
Out[17]: user_id         0
isbn          0
rating        2
dtype: int64
```

```
In [18]: users.isnull().sum()
```

```
Out[18]: user_id         0
Location          1
Age             110763
dtype: int64
```

```
In [19]: Ratings['rating'].fillna(Ratings.rating.mean(), inplace=True)
```

```
In [20]: Ratings.isnull().sum()
```

```
Out[20]: user_id         0
isbn          0
rating        0
dtype: int64
```

```
In [21]: users['Age'].fillna(users.Age.mean(), inplace=True)
```

```
In [22]: users.isnull().sum()
```

```
Out[22]: user_id         0
Location          1
Age              0
dtype: int64
```

```
In [23]: Ratings_with_name= Ratings.merge(Books, on='isbn')
```

```
In [24]: num_Ratings_df= Ratings_with_name.groupby('book_title').count()['rating'].reset_index
num_Ratings_df.rename(columns={'rating' : 'num_ratings'}, inplace=True)
```

```
In [25]: num_Ratings_df
```

```
Out[25]:
```

	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 Visita El Hospital (Clifford El Gran...	1
...
230232	Ã?Ã?l- Connection.	1
230233	Ã?Ã?lpiraten.	2
230234	Ã?Ã?rger mit Produkt X. Roman.	4
230235	Ã?Ã?stlich der Berge.	2
230236	Ã?Ã?thique en toc	2

230237 rows × 2 columns

```
In [26]: avg_Ratings_df= Ratings_with_name.groupby('book_title').mean()['rating'].reset_index(
avg_Ratings_df.rename(columns = {'rating':'avg_rating'}, inplace=True)
avg_Ratings_df
```

Out[26]:

	book_title	avg_rating
0	A Light in the Storm: The Civil War Diary of ...	2.25
1	Always Have Popsicles	0.00
2	Apple Magic (The Collector's series)	0.00
3	Beyond IBM: Leadership Marketing and Finance ...	0.00
4	Clifford Visita El Hospital (Clifford El Gran...	0.00
...
230232	Ã?Ã?l- Connection.	0.00
230233	Ã?Ã?lpiraten.	0.00
230234	Ã?Ã?rger mit Produkt X. Roman.	5.25
230235	Ã?Ã?stlich der Berge.	4.00
230236	Ã?Ã?thique en toc	4.00

230237 rows × 2 columns

```
In [27]: popular_df = num_Ratings_df.merge(avg_Ratings_df,on='book_title')
```

```
In [28]: popular_df
```

Out[28]:

	book_title	num_ratings	avg_rating
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 Visita El Hospital (Clifford El Gran...	1	0.00
...
230232	Ã?Ã?l- Connection.	1	0.00
230233	Ã?Ã?lpiraten.	2	0.00
230234	Ã?Ã?rger mit Produkt X. Roman.	4	5.25

230235	Ä?Ä?stlich der Berge.	2	4.00
230236	Ä?Ä?thique en toc	2	4.00

230237 rows × 3 columns

```
In [29]: popular_df= popular_df[popular_df['num_ratings']>=250].sort_values('avg_rating',ascen
popular_df
```

Out[29]:

	book_title	num_ratings	avg_rating
76833	Harry Potter and the Goblet of Fire (Book 4)	346	5.945087
76845	Harry Potter and the Prisoner of Azkaban (Book 3)	380	5.884211
76837	Harry Potter and the Order of the Phoenix (Boo...	316	5.544304
76825	Harry Potter and the Chamber of Secrets (Book 2)	501	5.177645
182901	The Hobbit : The Enchanting Prelude to The Lor...	261	4.969349

```
In [30]: popular_df.head()
```

Out[30]:

	book_title	num_ratings	avg_rating
76833	Harry Potter and the Goblet of Fire (Book 4)	346	5.945087
76845	Harry Potter and the Prisoner of Azkaban (Book 3)	380	5.884211
76837	Harry Potter and the Order of the Phoenix (Boo...	316	5.544304
76825	Harry Potter and the Chamber of Secrets (Book 2)	501	5.177645
182901	The Hobbit : The Enchanting Prelude to The Lor...	261	4.969349

```
In [31]: popular_df = popular_df.merge(Books,on='book_title').drop_duplicates('book_title')[['
```

```
In [32]: popular_df
```

Out[32]:

	book_title	book_author	num_ratings	avg_rating
0	Harry Potter and the Goblet of Fire (Book 4)	J. K. Rowling	346	5.945087
2	Harry Potter and the Prisoner of Azkaban (Book 3)	J. K. Rowling	380	5.884211
5	Harry Potter and the Order of the Phoenix (Boo...	J. K. Rowling	316	5.544304
9	Harry Potter and the Chamber of Secrets (Book 2)	J. K. Rowling	501	5.177645
12	The Hobbit : The Enchanting Prelude to The Lor...	J.R.R. TOLKIEN	261	4.969349

```
In [33]: Ratings_with_name
```

Out[33]:

	user_id	isbn	rating	book_title	book_author	year_of_publication	publisher
0	276725	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
1	2313	034545104X	5.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
2	6543	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
3	8680	034545104X	5.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
4	10314	034545104X	9.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
...
941136	250764	440106575	0.0	HIDDEN FIRES	JANETTE	1978	Dell

941137	250764	451157516	0.0	Cheyenne (Fortunes West, No 2)	A.R. Riefe	1988	New Amer Library (Mm)
941138	250764	048623715X	0.0	Glamorous Movie Stars of the Thirties: Paper D...	Tom Tierney	1982	Dover Publications
941139	250764	486256588	0.0	Schiaparelli Fashion Review: Paper Dolls in Fu...	Tom Tierney	1988	Dover Publications
941140	250764	515069434	0.0	Lady Laughing Eyes (To Have and to Hold)	Lee Damon	1984	Jove Books

941141 rows × 7 columns

In [34]: Ratings_with_name.head(10)

	user_id	isbn	rating	book_title	book_author	year_of_publication	publisher
0	276725	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
1	2313	034545104X	5.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
2	6543	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
3	8680	034545104X	5.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
4	10314	034545104X	9.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
5	23768	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
6	28266	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
7	28523	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
8	39002	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
9	50403	034545104X	9.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books

```
In [35]: x = Ratings_with_name.groupby('user_id').count()['rating'] > 200
smart_users = x[x].index
smart_users
```

```
Out[35]: Int64Index([ 254, 2276, 2766, 2977, 3363, 4017, 4385, 6251,
...
249111, 249628, 249862, 249894, 250184, 250405, 250764, 277427,
277639, 278418],
dtype='int64', name='user_id', length=737)
```

```
In [36]: filtered_rating = Ratings_with_name[Ratings_with_name['user_id'].isin(smart_users)]
filtered_rating
```

	user_id	isbn	rating	book_title	book_author	year_of_publication	publisher
2	6543	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
5	23768	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
7	28523	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
15	77940	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
16	81977	034545104X	0.0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books

...
941136	250764	440106575	0.0	HIDDEN FIRES	JANETTE RADCLIFFE	1978	Dell
941137	250764	451157516	0.0	Cheyenne (Fortunes West, No 2)	A.R. Riefe	1988	New Amer Library (Mm)
941138	250764	048623715X	0.0	Glamorous Movie Stars of the Thirties: Paper D...	Tom Tierney	1982	Dover Publications
941139	250764	486256588	0.0	Schiaparelli Fashion Review: Paper Dolls in Fu...	Tom Tierney	1988	Dover Publications
941140	250764	515069434	0.0	Lady Laughing Eyes (To Have and to Hold)	Lee Damon	1984	Jove Books

434830 rows × 7 columns

```
In [37]: Ratings_with_name.shape

Out[37]: (941141, 7)

In [38]: y = filtered_rating.groupby('book_title').count()['rating']>=50
famous_books = y[y].index

In [39]: final_Ratings = filtered_rating[filtered_rating['book_title'].isin(famous_books)]

In [40]: pt = final_Ratings.pivot_table(index='book_title',columns='user_id',values='rating')
pt
```

Out[40]:	user_id	254	2276	2766	2977	3363	4017	4385	6251	6323	6543	...	249111	249628	249862	249862
	book_title															
	1984	9.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
	1st to Die: A Novel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	9.0	...	NaN	0.0	NaN	NaN
	2nd Chance	NaN	10.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	...	NaN	NaN	NaN	NaN
	4 Blondes	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	NaN	...	NaN	NaN	NaN	NaN
	A Bend in the Road	0.0	NaN	7.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN

	Year of Wonders	NaN	NaN	NaN	7.0	NaN	NaN	NaN	NaN	NaN	0.0	...	NaN	NaN	NaN	NaN
	You Belong To Me	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	...	NaN	NaN	NaN	NaN
	Zen and the Art of Motorcycle Maintenance: An Inquiry into Values	NaN	NaN	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	...	NaN	NaN	NaN	NaN
	Zoya	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
	IO" Is for Outlaw"	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN

603 rows × 735 columns

```
In [41]: pt.fillna(0,inplace=True)
```

```
In [42]: pt
```

Out[42]:

	user_id	254	2276	2766	2977	3363	4017	4385	6251	6323	6543	...	249111	249628	249862	249
	book_title															
	1984	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
	1st to Die: A Novel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	...	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	
	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	
	A Bend in the Road	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
	
	Year of Wonders	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
	You Belong To Me	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	
	Zen and the Art of Motorcycle Maintenance: An Inquiry into Values	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	
	Zoya	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	
	"Is for Outlaw"	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	

603 rows × 735 columns

```
In [43]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [44]: similarity_scores = cosine_similarity(pt)
```

```
In [45]: similarity_scores.shape
```

Out[45]: (603, 603)

```
In [53]: def recommend(book_name):

    index = np.where(pt.index==book_name)[0][0]
    similar_items = sorted(list(enumerate(similarity_scores[index])),key=lambda x:x[1])

    data = []
    for i in similar_items:
        item = []
        temp_df = Books[Books['book_title'] == pt.index[i[0]]]
        item.extend(list(temp_df.drop_duplicates('book_title')['book_title'].values))
        item.extend(list(temp_df.drop_duplicates('book_title')['book_author'].values))
        item.extend(list(temp_df.drop_duplicates('book_title')['year_of_publication'].values))

        data.append(item)

    return data
```

```
In [54]: index = np.where(pt.index=='1984')[0][0]
index
```


Out[54]: 0

In [55]: `recommend('2nd Chance')`

Out[55]: `[['Four Blind Mice', 'James Patterson', '2002'],
['The Next Accident', 'LISA GARDNER', '2002'],
['The Murder Book', 'Jonathan Kellerman', '2003'],
['Cat & Mouse (Alex Cross Novels)', 'James Patterson', '1998']]`

In [56]: `recommend('1984')`

Out[56]: `[['Animal Farm', 'George Orwell', '2004'],
['Brave New World', 'Aldous Huxley', '1989'],
['The Vampire Lestat (Vampire Chronicles, Book II)', 'ANNE RICE', '1986'],
['The Handmaid's Tale', 'Margaret Atwood', '1989']]`

In [58]: `recommend('Harry Potter and the Chamber of Secrets (Book 2)')`

Out[58]: `[['Harry Potter and the Prisoner of Azkaban (Book 3)',
'J. K. Rowling',
'1999'],
['Harry Potter and the Goblet of Fire (Book 4)', 'J. K. Rowling', '2000'],
['Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))',
'J. K. Rowling',
'1999'],
['Harry Potter and the Sorcerer's Stone (Book 1)', 'J. K. Rowling', '1998']]`