

Recommendation System Assignment

Data Set : book

1. Import Necessary libraries

```
In [1]: import pandas as pd
import numpy as np

from matplotlib import pyplot as plt
import seaborn as sns

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

2. Import Data

```
In [2]: book_details = pd.read_csv('book.csv',encoding = 'latin1',index_col = 0)
book_details
```

```
Out[2]:
```

	UserID	BookTitle	BookRating
1	276726	Classical Mythology	5
2	276729	Clara Callan	3
3	276729	Decision in Normandy	6
4	276736	Flu: The Story of the Great Influenza Pandemic...	8
5	276737	The Mummies of Urunchi	6
...
9996	162121	American Fried: Adventures of a Happy Eater.	7
9997	162121	Cannibal In Manhattan	9
9998	162121	How to Flirt: A Practical Guide	7
9999	162121	Twilight	8
10000	162129	Kids Say the Darndest Things	6

10000 rows x 3 columns

3. Data Understanding

```
In [3]: book_details.head()
```

```
Out[3]:
```

	UserID	BookTitle	BookRating
1	276726	Classical Mythology	5
2	276729	Clara Callan	3
3	276729	Decision in Normandy	6
4	276736	Flu: The Story of the Great Influenza Pandemic...	8
5	276737	The Mummies of Urunchi	6

```
In [4]: book_details.shape
```

```
Out[4]: (10000, 3)
```

```
In [5]: book_details.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10000 entries, 1 to 10000
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
--  --
 0   User_ID     10000 non-null  int64
 1   Book_Title  10000 non-null  object
 2   Book_Rating 10000 non-null  int64
dtypes: int64(2), object(1)
memory usage: 312.5+ KB
```

```
In [6]: book_details.isna().sum()
```

```
Out[6]:
User_ID      0
Book_Title   0
Book_Rating  0
dtype: int64
```

```
In [7]: book_details.describe()
```

```
Out[7]:
```

	User_ID	Book_Rating
count	10000.000000	10000.00000
mean	95321.249800	7.56630
std	117645.703609	1.82152
min	8.000000	1.00000
25%	2103.000000	7.00000
50%	3757.000000	8.00000
75%	162052.000000	9.00000
max	278954.000000	10.00000

```
In [8]: book_details.dtypes
```

```
Out[8]:
User_ID      int64
Book_Title   object
Book_Rating  int64
dtype: object
```

```
In [9]: book_details.nunique()
```

```
Out[9]:
User_ID      2182
Book_Title   9659
Book_Rating  10
dtype: int64
```

4. User Based Colloborative filtering

```
In [10]: book_details.columns = ["UserID","BookTitle","BookRating"]
book_details
```

```
Out[10]:
```

	UserID	BookTitle	BookRating
1	276726	Classical Mythology	5
2	276729	Clara Callan	3
3	276729	Decision in Normandy	6
4	276736	Flu: The Story of the Great Influenza Pandemic...	8
5	276737	The Mummies of Urunchi	6
...
9996	162121	American Fried: Adventures of a Happy Eater.	7
9997	162121	Cannibal In Manhattan	9
9998	162121	How to Flirt: A Practical Guide	7
9999	162121	Twilight	8
10000	162129	Kids Say the Darndest Things	6

10000 rows x 3 columns

```
In [11]: # Sort by User IDs
```

```
book_details.sort_values(['UserID'])
```

```
Out[11]:
```

	UserID	BookTitle	BookRating
2402	8	Wings	5
2401	8	The Western way: A practical guide to the West...	5
2400	8	Ancient Celtic Romances	5
2403	8	Truckers	5
2406	8	The Art Of Celia	7
...
2396	278954	La cr�nica del Per� (Cr�nicas de Am�rica)	7
2399	278954	Celtic Mythology (Library of the World's Myths...	8
2394	278954	A corrente de vav�s Scott	7
2395	278954	As walk�rias	7
2398	278954	A Treasury of Irish Myth, Legend, and Folklore	6

10000 rows x 3 columns

```
In [12]: # number of unique users in the dataset
```

```
len(book_details['UserID'].unique())
```

```
Out[12]: 2182
```

```
In [13]: # number of unique books in the dataset
```

```
len(book_details['BookTitle'].unique())
```

```
Out[13]: 9659
```

```
In [14]: # number of unique books in the dataset
```

```
len(book_details['BookRating'].unique())
```

```
Out[14]: 10
```

```
In [15]: book_details.loc[book_details['BookRating'] == 'small', 'BookRating'] = 0
book_details.loc[book_details['BookRating'] == 'large', 'BookRating'] = 1
```

```
In [16]: book_details.BookRating.value_counts()
```

```
Out[16]:
8      2283
7      2676
10     1732
9      1493
5      1807
6       920
4       237
3       146
2        63
1        43
Name: BookRating, dtype: int64
```

```
In [17]: plt.figure(figsize = (20,6))
```

```
sns.distplot(book_details.BookRating)
plt.show()
```



5. Correlation Matrix

```
In [18]: # converting long data into wide data using pivot table
```

```
book_details_corr = book_details.pivot_table(index='UserID',columns = 'BookTitle',values = 'BookRating').reset_index(drop = True)
```

```
Out[18]:
```

BookTitle	Jason, Madison & Jason, Madison & Stories:Merrill1985,McClelland & Other PC Drives & Repairing PC Drives & '48 'O Au No Keia: Voices from Hawai'i's Mahu and Transgender Communities	...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR	01-01-00: A Novel of the Millennium	1,401 More Things That P-Ss Me Off	Commandments Of Dating	10 Great Fantasy Short Stories	100 Great Fantasy Short Stories	Zora Hurston and the Chinaberry Tree (Reading Rainbow Book)	Even Monkeys Fall from Trees! and Other Japanese Proverbs	U Won't Learn from Yout And Other Creative Maladjustment	More More, Said the Baby	IO! Is for Outlaw	!Surely You're Joking, Mr. Feynman!: Adventures of a Curious Character	Well, there's your problem!: Cartoons	Para l C
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
2177	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2178	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2179	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2180	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2181	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

2182 rows x 9659 columns

```
In [19]: # Replacing the index values by unique user Ids
```

```
book_details_corr.index = book_details['UserID'].unique()
book_details_corr
```

```
Out[19]:
```

BookTitle	Jason, Madison & Jason, Madison & Stories:Merrill1985,McClelland & Other PC Drives & Repairing PC Drives & '48 'O Au No Keia: Voices from Hawai'i's Mahu and Transgender Communities	...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR	01-01-00: A Novel of the Millennium	1,401 More Things That P-Ss Me Off	Commandments Of Dating	10 Great Fantasy Short Stories	100 Great Fantasy Short Stories	Zora Hurston and the Chinaberry Tree (Reading Rainbow Book)	Even Monkeys Fall from Trees! and Other Japanese Proverbs	U Won't Learn from Yout And Other Creative Maladjustment	More More, Said the Baby	IO! Is for Outlaw	!Surely You're Joking, Mr. Feynman!: Adventures of a Curious Character	Well, there's your problem!: Cartoons	Para l C
276726	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
276729	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
276736	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
276737	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
276744	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
162107	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
162109	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
162113	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
162121	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
162129	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

2182 rows x 9659 columns

```
In [20]: # Impute those NaNs with 0 values
```

```
book_details_corr.fillna(0,inplace = True)
book_details_corr
```

```
Out[20]:
```

BookTitle	Jason, Madison & Jason, Madison & Stories:Merrill1985,McClelland & Other PC Drives & Repairing PC Drives & '48 'O Au No Keia: Voices from Hawai'i's Mahu and Transgender Communities	...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR	01-01-00: A Novel of the Millennium	1,401 More Things That P-Ss Me Off	Commandments Of Dating	10 Great Fantasy Short Stories	100 Great Fantasy Short Stories	Zora Hurston and the Chinaberry Tree (Reading Rainbow Book)	Even Monkeys Fall from Trees! and Other Japanese Proverbs	U Won't Learn from Yout And Other Creative Maladjustment	More More, Said the Baby	IO! Is for Outlaw	!Surely You're Joking, Mr. Feynman!: Adventures of a Curious Character	Well, there's your problem!: Cartoons	Para l C
276726	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
276729	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
276736	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
276737	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
276744	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
...
162107	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
162109	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
162113	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
162121	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
162129	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

2182 rows x 9659 columns

6. Calculating Cosine Similarity between Users on Array data

```
In [21]: from sklearn.metrics import pairwise_distances
```

```
from scipy.spatial.distance import cosine, correlation
```

```
In [22]: book_sim_1 = 1 - pairwise_distances(book_details_corr.values,metric = 'cosine')
```

```
book_sim_1
```

```
Out[22]: array([[1., 0., 0., ..., 0., 0., 0., 0.],
```

```
[0., 1., 0., ..., 0., 0., 0., 0.],
```

```
[0., 0., 1., ..., 0., 0., 0., 0.],
```

```
...,
```

```
[0., 0., 0., ..., 1., 0., 0.],
```

```
[0., 0., 0., ..., 0., 1., 0.],
```

```
[0., 0., 0., ..., 0., 0., 1.]])
```

```
In [23]: # Store the results in a dataframe format
```

```
book_sim_2 = pd.DataFrame(book_sim_1)
book_sim_2
```

```
Out[23]:
```

	0	1	2	3	4	5	6	7	8	9	...	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181
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	0.0	0.0	0.0	0.0	0.0	0.0
1	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	0.0	0.0	0.0	0.0	0.0
2	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	0.0	0.0	0.0	0.0
3	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	0.0	0.0	0.0
4	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	0.0	0.0
...
2177	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	1.0	0.0	0.0	0.0
2178	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	1.0	0.0	0.0
2179	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	1.0	0.0
2180	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	1.0
2181	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

2182 rows x 2182 columns

```
In [24]: # Set the index and column names to user ids
```

```
book_sim_2.index = book_details['UserID'].unique()
book_sim_2.columns = book_details['UserID'].unique()
book_sim_2
```

```
Out[24]:
```

	276726	276729	276736	276737	276744	276745	276747	276748	276751	276754	...	162085	162091	
--	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	-----	--------	--------	--