

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

```
In [2]: books = pd.read_csv("assbook.csv",encoding='Latin1')
```

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

Out [3]:

	Unnamed: 0	User.ID	Book.Title	Book.Rating
0	1	276726	Classical Mythology	5
1	2	276729	Clara Callan	3
2	3	276729	Decision in Normandy	6
3	4	276736	Flu: The Story of the Great Influenza Pandemic...	8
4	5	276737	The Mummies of Urumchi	6

```
In [4]: books_df=books.iloc[:,1:]
books_df
```

Out [4]:

	User.ID	Book.Title	Book.Rating
0	276726	Classical Mythology	5
1	276729	Clara Callan	3
2	276729	Decision in Normandy	6
3	276736	Flu: The Story of the Great Influenza Pandemic...	8
4	276737	The Mummies of Urumchi	6
...	...	...	...
9995	162121	American Fried: Adventures of a Happy Eater.	7
9996	162121	Cannibal In Manhattan	9
9997	162121	How to Flirt: A Practical Guide	7
9998	162121	Twilight	8
9999	162129	Kids Say the Darndest Things	6

10000 rows × 3 columns

```
In [5]: books_df.sort_values('User.ID')
```

```
Out[5]:
```

	User.ID	Book.Title	Book.Rating
2401	8	Wings	5
2400	8	The Western way: A practical guide to the West...	5
2399	8	Ancient Celtic Romances	5
2402	8	Truckers	5
2405	8	The Art Of Celtia	7
...	...	...	...
2395	278854	La cr��nica del Per�� (Cr��nicas de Am��rica)	7
2398	278854	Celtic Mythology (Library of the World's Myths...	8
2393	278854	A corrente de Trewis Scott	7
2394	278854	As valk��rias	7
2397	278854	A Treasury of Irish Myth, Legend, and Folklore	6

10000 rows × 3 columns

```
In [6]: len(books_df['User.ID'].unique())
```

```
Out[6]: 2182
```

```
In [7]: len(books_df['Book.Title'].unique())
```

```
Out[7]: 9659
```

```
In [8]: books_df['Book.Title'].value_counts()
```

```
Out[8]: Fahrenheit 451                    5
Charlie and the Chocolate Factory         4
The Subtle Knife (His Dark Materials, Book 2) 4
Vanished                                 4
Ender's Game (Ender Wiggins Saga (Paperback)) 4
..
Murder on St. Mark's Place (Gaslight Mysteries) 1
State of Grace                            1
Valsalva's Maneuver: Mots Justes and Indispensable Terms 1
I love you, I hate you                    1
Kids Say the Darndest Things              1
Name: Book.Title, Length: 9659, dtype: int64
```

```
In [9]: user_books_df = books_df.pivot_table(index='User.ID', columns='Book.Title', values='Rating')
user_books_df
```

Out [9]:

Book.Title	Jason, Madison & the	Stories; Merrill; 1985; McClelland & Co.	Other	Repairing PC Drives & Software	'48	'O Au No Keia: Voices from Hawai'i's Mahu and Transgender Communities	...ANI TH HORS HE ROD IN ON TH PEOPLES KENNETH STARR
User.ID							
8	NaN		NaN	NaN	NaN	NaN	NaN
9	NaN		NaN	NaN	NaN	NaN	NaN
10	NaN		NaN	NaN	NaN	NaN	NaN
12	NaN		NaN	NaN	NaN	NaN	NaN
14	NaN		NaN	NaN	NaN	NaN	NaN
...	...		...	...	...	...	...
278846	NaN		NaN	NaN	NaN	NaN	NaN
278849	NaN		NaN	NaN	NaN	NaN	NaN
278851	NaN		NaN	NaN	NaN	NaN	NaN
278852	NaN		NaN	NaN	NaN	NaN	NaN
278854	NaN		NaN	NaN	NaN	NaN	NaN

2182 rows × 9659 columns

```
In [10]: # Converting NaN into 0
user_books_df.fillna(0,inplace=True)
user_books_df
```

Out[10]:

Book.Title	Jason, Madison &	Stories;Merril;1985;McClelland &	Other	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
User.ID							
8	0.0		0.0	0.0	0.0	0.0	0.0
9	0.0		0.0	0.0	0.0	0.0	0.0
10	0.0		0.0	0.0	0.0	0.0	0.0
12	0.0		0.0	0.0	0.0	0.0	0.0
14	0.0		0.0	0.0	0.0	0.0	0.0
...	...		...	...	...	...	...
278846	0.0		0.0	0.0	0.0	0.0	0.0
278849	0.0		0.0	0.0	0.0	0.0	0.0
278851	0.0		0.0	0.0	0.0	0.0	0.0
278852	0.0		0.0	0.0	0.0	0.0	0.0
278854	0.0		0.0	0.0	0.0	0.0	0.0

2182 rows × 9659 columns

```
In [11]: # Calculating Cosine Similarity between Users
from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import cosine,correlation
```

```
In [12]: user_sim = 1-pairwise_distances(user_books_df.values,metric='cosine')
user_sim
```

```
Out[12]: array([[1., 0., 0., ..., 0., 0., 0.],
                [0., 1., 0., ..., 0., 0., 0.],
                [0., 0., 1., ..., 0., 0., 0.],
                ...,
                [0., 0., 0., ..., 1., 0., 0.],
                [0., 0., 0., ..., 0., 1., 0.],
                [0., 0., 0., ..., 0., 0., 1.]])
```

```
In [13]: user_sim_df = pd.DataFrame(user_sim)
```

```
In [14]: # Set the index and column names to user ids
user_sim_df.index=books_df['User.ID'].unique()
user_sim_df.columns=books_df['User.ID'].unique()
```

```
In [15]: np.fill_diagonal(user_sim,0)
user_sim_df.iloc[0:5,0:5]
```

Out[15]:

	276726	276729	276736	276737	276744
276726	0.0	0.0	0.0	0.0	0.0
276729	0.0	0.0	0.0	0.0	0.0
276736	0.0	0.0	0.0	0.0	0.0
276737	0.0	0.0	0.0	0.0	0.0
276744	0.0	0.0	0.0	0.0	0.0

```
In [16]: # Finding Similar Users
user_sim_df.idxmax(axis=1)
```

Out[16]:

276726	276726
276729	276726
276736	276726
276737	276726
276744	276726
	...
162107	276726
162109	276726
162113	161453
162121	276726
162129	276726

Length: 2182, dtype: int64

```
In [17]: books_df[(books_df['User.ID']==162121)|(books_df['User.ID']==276726)
```

```
Out[17]:
```

	User.ID	Book.Title	Book.Rating
0	276726	Classical Mythology	5
9990	162121	The Cloister Walk	7
9991	162121	Open Water	5
9992	162121	The Evolution of Jane	8
9993	162121	AT PARADISE GATE	8
9994	162121	I Should Have Stayed Home: The Worst Trips of ...	8
9995	162121	American Fried: Adventures of a Happy Eater.	7
9996	162121	Cannibal In Manhattan	9
9997	162121	How to Flirt: A Practical Guide	7
9998	162121	Twilight	8

```
In [18]: user_1 = books_df[(books_df['User.ID']==162121)]
user_2 =books_df[(books_df['User.ID']==276726)]
user_1['Book.Title']
```

```
Out[18]: 9990          The Cloister Walk
9991          Open Water
9992          The Evolution of Jane
9993          AT PARADISE GATE
9994    I Should Have Stayed Home: The Worst Trips of ...
9995          American Fried: Adventures of a Happy Eater.
9996          Cannibal In Manhattan
9997          How to Flirt: A Practical Guide
9998          Twilight
Name: Book.Title, dtype: object
```

```
In [19]: user_2['Book.Title']
```

```
Out[19]: 0    Classical Mythology
Name: Book.Title, dtype: object
```

```
In [20]: pd.merge(user_1,user_2,on='Book.Title',how='outer')
```

```
Out[20]:
```

	User.ID_x	Book.Title	Book.Rating_x	User.ID_y	Book.Rating_y
0	162121.0	The Cloister Walk	7.0	NaN	NaN
1	162121.0	Open Water	5.0	NaN	NaN
2	162121.0	The Evolution of Jane	8.0	NaN	NaN
3	162121.0	AT PARADISE GATE	8.0	NaN	NaN
4	162121.0	I Should Have Stayed Home: The Worst Trips of ...	8.0	NaN	NaN
5	162121.0	American Fried: Adventures of a Happy Eater.	7.0	NaN	NaN
6	162121.0	Cannibal In Manhattan	9.0	NaN	NaN
7	162121.0	How to Flirt: A Practical Guide	7.0	NaN	NaN
8	162121.0	Twilight	8.0	NaN	NaN
9	NaN	Classical Mythology	NaN	276726.0	5.0

```
In [ ]:
```