

```
In [1]: import pandas as pd
import numpy as np
from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import cosine, correlation
```

```
In [3]: books=pd.read_csv('book.csv', encoding = 'Latin1')
books
```

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
...	...	...	...	...
9995	9996	162121	American Fried: Adventures of a Happy Eater.	7
9996	9997	162121	Cannibal In Manhattan	9
9997	9998	162121	How to Flirt: A Practical Guide	7
9998	9999	162121	Twilight	8
9999	10000	162129	Kids Say the Darndest Things	6

10000 rows × 4 columns

```
In [5]: books.shape
```

Out[5]: (10000, 4)

```
In [7]: books.head()
```

Out[7]:

	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 [8]: books2=books.iloc[:,1:]
books2
```

Out[8]:

	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 [9]: books2.sort_values(['User.ID'])
```

Out[9]:

	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 [10]: len(books2['User.ID'].unique())
```

Out[10]: 2182

```
In [11]: len(books2['Book.Title'].unique())
```

Out[11]: 9659

```
In [12]: books3=books2.pivot_table(index='User.ID',columns='Book.Title',values='Book.Rating',aggfunc='mean')
books3
```

Out[12]:

	Book.Title	Jason, Madison &amp;	Stories;Merril;1985;McClelland &amp;	Other	Repairing PC Drives &amp;	'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	01-1 A No Miller
	0	NaN		NaN	NaN	NaN	NaN	NaN	
	1	NaN		NaN	NaN	NaN	NaN	NaN	
	2	NaN		NaN	NaN	NaN	NaN	NaN	
	3	NaN		NaN	NaN	NaN	NaN	NaN	
	4	NaN		NaN	NaN	NaN	NaN	NaN	
	...	...		...	...	...	...	...	
	2177	NaN		NaN	NaN	NaN	NaN	NaN	
	2178	NaN		NaN	NaN	NaN	NaN	NaN	
	2179	NaN		NaN	NaN	NaN	NaN	NaN	
	2180	NaN		NaN	NaN	NaN	NaN	NaN	
	2181	NaN		NaN	NaN	NaN	NaN	NaN	

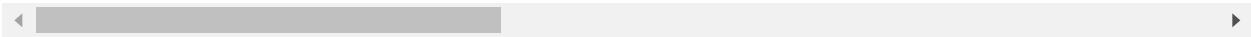
2182 rows × 9659 columns

```
In [13]: books3.index=books2['User.ID'].unique()  
books3
```

Out[13]:

Book.Title	Jason, Madison &amp;	Stories;Merril;1985;McClelland &amp;	Other Repairing PC Drives &amp;	'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	01-1 A No Miller
276726	NaN		NaN	NaN	NaN	NaN	
276729	NaN		NaN	NaN	NaN	NaN	
276736	NaN		NaN	NaN	NaN	NaN	
276737	NaN		NaN	NaN	NaN	NaN	
276744	NaN		NaN	NaN	NaN	NaN	
...	...		...	...	...	...	
162107	NaN		NaN	NaN	NaN	NaN	
162109	NaN		NaN	NaN	NaN	NaN	
162113	NaN		NaN	NaN	NaN	NaN	
162121	NaN		NaN	NaN	NaN	NaN	
162129	NaN		NaN	NaN	NaN	NaN	

2182 rows × 9659 columns

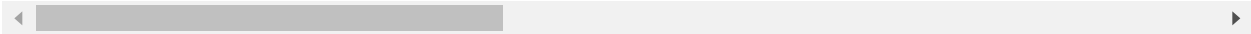


```
In [14]: books3.fillna(0,inplace=True)
books3
```

Out[14]:

Book.Title	Jason, Madison &amp;	Stories;Merril;1985;McClelland &amp;	Other Repairing PC Drives &amp;	'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	01-0' A Nov Millenr
276726	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
276736	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
276744	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
162109	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
162121	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

2182 rows × 9659 columns



# Calculating Cosine Similarity between Users on array data

```
In [15]: user_sim=1-pairwise_distances(books3.values,metric='cosine')
user_sim
```

```
Out[15]: 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.]])
```

## Store the results in a dataframe format

```
In [16]: user_sim2=pd.DataFrame(user_sim)
user_sim2
```

```
Out[16]:
```

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

2182 rows × 2182 columns



## Set the index and column names to user ids

```
In [17]: user_sim2.index=books2['User.ID'].unique()
user_sim2.columns=books2['User.ID'].unique()
user_sim2
```

Out[17]:

	276726	276729	276736	276737	276744	276745	276747	276748	276751	276754	...	1
276726	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	
276729	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	
276736	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	
276737	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	
276744	0.0	0.0	0.0	0.0	1.0	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	...	
162109	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	...	
162121	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	...	

2182 rows × 2182 columns



Nullifying diagonal values

```
In [18]: np.fill_diagonal(user_sim,0)
user_sim2
```

Out[18]:

	276726	276729	276736	276737	276744	276745	276747	276748	276751	276754	...	1
276726	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	...	
276736	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	...	
276744	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	...	
162109	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	...	
162121	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	...	

2182 rows × 2182 columns



## Most Similar Users

```
In [19]: user_sim2.idxmax(axis=1)
```

Out[19]:

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

Length: 2182, dtype: int64

## extract the books which userId 162107 & 276726 have watched

```
In [20]: books2[(books2['User.ID']==162107) | (books2['User.ID']==276726)]
```

Out[20]:

	User.ID	Book.Title	Book.Rating
0	276726	Classical Mythology	5
9987	162107	What's Bred in the Bone	7



**extract the books which userId 276729 & 276726 have watched**

```
In [21]: books2[(books2['User.ID']==276729) | (books2['User.ID']==276726)]
```

```
Out[21]:
```

	User.ID	Book.Title	Book.Rating
0	276726	Classical Mythology	5
1	276729	Clara Callan	3
2	276729	Decision in Normandy	6

```
In [22]: user_1=books2[(books2['User.ID']==276729)]
user_2=books2[(books2['User.ID']==276726)]
```

```
In [23]: user_1['Book.Title']
```

```
Out[23]: 1          Clara Callan
2    Decision in Normandy
Name: Book.Title, dtype: object
```

```
In [24]: user_2['Book.Title']
```

```
Out[24]: 0    Classical Mythology
Name: Book.Title, dtype: object
```

```
In [25]: pd.merge(user_1,user_2,on='Book.Title',how='outer')
```

```
Out[25]:
```

	User.ID_x	Book.Title	Book.Rating_x	User.ID_y	Book.Rating_y
0	276729.0	Clara Callan	3.0	NaN	NaN
1	276729.0	Decision in Normandy	6.0	NaN	NaN
2	NaN	Classical Mythology	NaN	276726.0	5.0

```
In [ ]:
```