1. Import libraries

In [1]:

```python
import pandas as pd
import numpy as np
from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import cosine,correlation
```

2. Import Data

In [2]:

```python
books_data = pd.read_csv('book.csv')
books_data
```

Out[2]:

	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

3. Data Understanding

In [14]:

```python
books = books_data.iloc[:,1:]
books
```

Out[14]:

	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 [15]:

```python
books.sort_values(['User.ID'])
```

Out[15]:

	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 crónica del Perú (Crónicas de América)	7
2398	278854	Celtic Mythology (Library of the World's Myths...	8
2393	278854	A corrente de Trewis Scott	7
2394	278854	As valkírias	7
2397	278854	A Treasury of Irish Myth, Legend, and Folklore	6

10000 rows × 3 columns

In [16]:

```python
books.shape
```

Out[16]:

```
(10000, 3)
```

In [18]:

```python
len(books['User.ID'].unique())
```

Out[18]:

```
2182
```

In [20]:

```python
len(books['Book.Title'].unique())
```

Out[20]:

```
9659
```

In [24]:

```
book_data_3 = books.pivot_table( values='Book.Rating', index='User.ID',columns='Book.Title'
book_data_3
```

Out[24]:

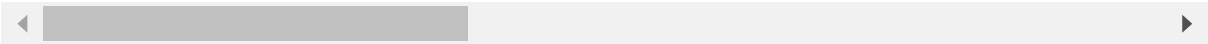| Book.Title	Jason, Madison &amp	Other Stories;Merril;1985;McClelland &amp	Repairing PC Drives &amp	'48	'O Au No Keia: Voices from Hawai'I's Mahu and Transgender Communities	...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR
User.ID						
8	NaN	NaN	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN	NaN	NaN
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
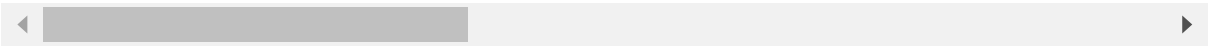
In [29]:

```python
# Replacing the index values by unique user Ids
book_data_3.index = books['User.ID'].unique()
book_data_3
```

Out[29]:

| Book.Title	Jason, Madison &amp	Stories;Merril;1985;McClelland &amp	Other PC Drives &amp	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	
276726	NaN	NaN	NaN	NaN	NaN	NaN	
276729	NaN	NaN	NaN	NaN	NaN	NaN	
276736	NaN	NaN	NaN	NaN	NaN	NaN	
276737	NaN	NaN	NaN	NaN	NaN	NaN	
276744	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	
162107	NaN	NaN	NaN	NaN	NaN	NaN	
162109	NaN	NaN	NaN	NaN	NaN	NaN	
162113	NaN	NaN	NaN	NaN	NaN	NaN	
162121	NaN	NaN	NaN	NaN	NaN	NaN	
162129	NaN	NaN	NaN	NaN	NaN	NaN	

2182 rows × 9659 columns

In [31]:

```python
# Impute those NoNs with 0 values
book_data_3.fillna(0, inplace=True,)
book_data_3
```

Out[31]:

Book.Title	Jason, Madison &amp	Other Stories;Merril;1985;McClelland &amp	Repairing PC Drives &amp	'48	'O Au No Keia: Voices from Hawai'l's Mahu and Transgender Communities	...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR	O A Mil
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
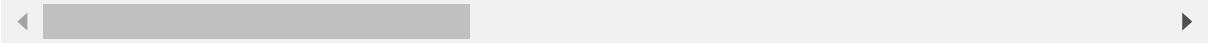
To Calculate Cosine Similarity between Users on array data

In [33]:

```python
user_sim = pairwise_distances(book_data_3.values,metric='cosine')
user_sim
```

Out[33]:

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

In [35]:

```python
user_sim = 1 - pairwise_distances(book_data_3.values,metric='cosine')
user_sim
```

Out[35]:

```
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 [37]:

```python
# Store the results in a dataframe format
user_sim_2 = pd.DataFrame(user_sim)
user_sim_2
```

Out[37]:

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

2182 rows × 2182 columns

In [38]:

```python
# Set the index and columns names to user ids
user_sim_2.index = books['User.ID'].unique()
user_sim_2.columns = books['User.ID'].unique()
user_sim_2
```

Out[38]:

	276726	276729	276736	276737	276744	276745	276747	276748	276751	276754	..
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 digonal values

In [39]:

```python
np.fill_diagonal(user_sim,0)
user_sim_2
```

Out[39]:

	276726	276729	276736	276737	276744	276745	276747	276748	276751	276754	..
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

In [40]:

```python
user_sim_2.idxmax(axis = 0)
```

Out[40]:

```
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 [52]:

```python
books[(books['User.ID']==162109) | (books['User.ID']==276726)]
```

Out[52]:

	User.ID	Book.Title	Book.Rating
0	276726	Classical Mythology	5
9988	162109	The Flower in the Skull	10

In [53]:

```python
books[(books['User.ID']==276726) | (books['User.ID']==276726)]
```

Out[53]:

	User.ID	Book.Title	Book.Rating
0	276726	Classical Mythology	5

In [54]:

```python
books[(books['User.ID']==276737) | (books['User.ID']==276726)]
```

Out[54]:

	User.ID	Book.Title	Book.Rating
0	276726	Classical Mythology	5
4	276737	The Mummies of Urumchi	6

In [70]:

```python
user_1 = books[(books['User.ID']==276726)]
user_2 = books[(books['User.ID']==276744)]
```

In [71]:

```python
user_1['Book.Title']
```

Out[71]:

```
0    Classical Mythology
Name: Book.Title, dtype: object
```

In [72]:

```python
user_2['Book.Title']
```

Out[72]:

```
5    The Kitchen God's Wife
Name: Book.Title, dtype: object
```

In [73]:

```python
pd.merge(user_1,user_2,on='Book.Title',how='outer')
```

Out[73]:

	User.ID_x	Book.Title	Book.Rating_x	User.ID_y	Book.Rating_y
0	276726.0	Classical Mythology	5.0	NaN	NaN
1	NaN	The Kitchen God's Wife	NaN	276744.0	7.0

In [76]:

```python
user_3 = books[(books['User.ID']==276736)]
user_4 = books[(books['User.ID']==276744)]
```

In [77]:

```python
user_3['Book.Title']
```

Out[77]:

```
3    Flu: The Story of the Great Influenza Pandemic...
Name: Book.Title, dtype: object
```

In [78]:

```python
user_4['Book.Title']
```

Out[78]:

```
5    The Kitchen God's Wife
Name: Book.Title, dtype: object
```

In [79]:

```python
pd.merge(user_3,user_4,on='Book.Title',how='outer')
```

Out[79]:

	User.ID_x	Book.Title	Book.Rating_x	User.ID_y	Book.Rating_y
0	276736.0	Flu: The Story of the Great Influenza Pandemic...	8.0	NaN	NaN
1	NaN	The Kitchen God's Wife	NaN	276744.0	7.0

In [ ]: