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

In [3]:
```
book=pd.read_csv('book..csv',encoding='latin-1')
book
```

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 [29]:
```
if 'Unnamed: 0' in df.columns:
    print("Column 'Unnamed: 0' exists.")
else:
    print("Column 'Unnamed: 0' does not exist.")
```

```
Column 'Unnamed: 0' does not exist.
```

In [30]:
```
df =book.drop(['Unnamed: 0'],axis=1)
```

In [31]:
```
df =df.rename({'User.ID':'user_id','Book.Title':'book_title','Book.Rating':'book_rating'},axis=1)
```

In [32]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
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: 234.5+ KB
```

In [33]:
```
len(df.user_id.unique())
```

Out[33]:
```
2182
```

In [34]:
```
len(df.book_title.unique())
```

Out[34]:
```
9659
```

In [35]:
```
df1 = df.drop_duplicates(['user_id','book_title'])
```

In [36]:
```
books = df1.pivot(index='user_id',columns='book_title',values='book_rating').reset_index(drop=True)
```

In [37]:
```
books
```

Out[37]:

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	01-01-00: A Novel of the Millennium	1,401 More Things That P*Ss Me Off	10 Commandments Of Dating	100 Great Fantasy Short, Short Stories	.
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
...	...	...	...	...	...	...	...	...	...	...	.
2177	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
2178	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
2179	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
2180	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
2181	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.

2182 rows × 9659 columns

In [38]:
```
books.index = df.user_id.unique()
```

In [39]:
```
books
```

Out[39]:

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	01-01-00: A Novel of the Millennium	1,401 More Things That P*Ss Me Off	10 Commandments Of Dating	100 Great Fantasy Short, Short Stories	.
276726	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
276729	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
276736	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
276737	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
276744	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
...	...	...	...	...	...	...	...	...	...	...	.
162107	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
162109	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
162113	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
162121	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.
162129	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	.

2182 rows × 9659 columns

In [40]:
```
books.fillna(0, inplace=True)
```

In [41]:
```
books
```

Out[41]:

book_title	Jason, Madison &amp	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	01-01-00: A Novel of the Millennium	1,401 More Things That P*Ss Me Off	10 Commandments Of Dating	100 Great Fantasy Short, Short Stories	...
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 × 9659 columns

In [42]:
```
from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import cosine, correlation
```

In [43]:
```
df2 = 1 - pairwise_distances( books.values,metric='cosine')
```

In [44]:
```
df2
```

Out[44]:
```
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 [45]:
```
books2 = pd.DataFrame(df2)
```

In [46]:
```
books2.index = df1.user_id.unique()
books2.columns = df1.user_id.unique()
```

In [47]:
```
books2.iloc[0:5, 0:5]
```

Out[47]:

	276726	276729	276736	276737	276744
276726	1.0	0.0	0.0	0.0	0.0
276729	0.0	1.0	0.0	0.0	0.0
276736	0.0	0.0	1.0	0.0	0.0
276737	0.0	0.0	0.0	1.0	0.0
276744	0.0	0.0	0.0	0.0	1.0

In [48]:
```
np.fill_diagonal(df2, 0)
books2.iloc[0:5, 0:5]
```

Out[48]:

	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 [49]:
```
books2.idxmax(axis=1)[0:5]
```

Out[49]:
```
276726    276726
276729    276726
276736    276726
276737    276726
276744    276726
dtype: int64
```

In [ ]:

In [ ]: