Recommandation System Q1 Books

Problem statement.

Build a recommender system by using cosine simillarties score.

1. Import Libs

In [26]:

```python
import pandas as pd
import numpy as np
from sklearn.metrics import pairwise_distances
```

2. Import Data

In [7]:

```python
book_data = pd.read_csv('book.csv',encoding='Latin1')
book_data
```

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
...	...	...	...	...
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

Droping (Unnamed: 0) column

In [13]:

```python
book_data2 = book_data.drop(['Unnamed: 0'],axis=1)
```

In [14]:

```python
book_data2
```

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

Sort values by User IDs

In [16]:

```python
book_data2.sort_values(['User.ID'])
```

Out[16]:

	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 crÃ³nica del PerÃº (CrÃ³nicas de AmÃ©rica)	7
2398	278854	Celtic Mythology (Library of the World's Myths...	8
2393	278854	A corrente de Trewis Scott	7
2394	278854	As valkÃrias	7
2397	278854	A Treasury of Irish Myth, Legend, and Folklore	6

10000 rows × 3 columns

In [17]:

```python
book_data2.shape
```

Out[17]:

```
(10000, 3)
```

Total number of unique users in the dataset

In [21]:

```python
book_data2['User.ID'].nunique()
```

Out[21]:

```
2182
```

Total number of unique books in the dataset

In [22]:

```python
book_data2['Book.Title'].nunique()
```

Out[22]:

```
9659
```

4. Data Prearation

Using pivot table to replace long data into wide data

In [24]:

```
user_pivot_table = book_data2.pivot_table(values='Book.Rating', index='User.ID',columns='Bo
user_pivot_table.columns = book_data2['Book.Title'].unique()
user_pivot_table
```

Out[24]:

User.ID	Classical Mythology	Clara Callan	Decision in Normandy	Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It	The Mummies of Urumchi	The Kitchen God's Wife	What If?: The World's Foremost Military Historians Imagine What Might Have Been	PLEADING GUILTY	
8	0.0	0.0	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	0.0	0.0	
10	0.0	0.0	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	0.0	0.0	
14	0.0	0.0	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	0.0	0.0	
278849	0.0	0.0	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	0.0	0.0	
278852	0.0	0.0	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	0.0	0.0	

2182 rows × 9659 columns

5. Model Building

Metric = Cosine

Calculating Cosine Similarity between Users on array data

In [27]:

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

Out[27]:

```
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.]])
```

Set the index and column names to user ids

In [38]:

```
cosine_metrics_user = pd.DataFrame(user_sim)
cosine_metrics_user.index=book_data2['User.ID'].unique()
cosine_metrics_user.columns=book_data2['User.ID'].unique()
cosine_metrics_user
```

Out[38]:

	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

Most Similar Users

In [39]:

```
cosine_metrics_user.idxmax(axis=1)
```

Out[39]:

```
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 user Id 276729 & 276726 have watched

In [40]:

```
book_data2[(book_data2['User.ID']==276729) | (book_data2['User.ID']==276726)]
```

Out[40]:

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

Extract the books which user Id 162107 & 276726 have watched

In [41]:

```
book_data2[(book_data2['User.ID']==162107) | (book_data2['User.ID']==276726)]
```

Out[41]:

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

Extract the books which user Id 8,9 & 10 have watched

In [44]:

```python
book_data2[(book_data2['User.ID']==8) | (book_data2['User.ID']==9) | (book_data2['User.ID'
```

Out[44]:

	User.ID	Book.Title	Book.Rating
2399	8	Ancient Celtic Romances	5
2400	8	The Western way: A practical guide to the West...	5
2401	8	Wings	5
2402	8	Truckers	5
2403	8	Keepers of the Earth Teachers Guide	6
2404	8	The Celts Activity Book	6
2405	8	The Art Of Celtia	7
2406	9	The Book of Kells: Selected Plates in Full Color	6
2407	10	SINGLE & SINGLE	6

END