In [2]:
```
import numpy as np
import pandas as pd
```

In [3]:
```
books = pd.read_csv(r'C:\Users\vibha\OneDrive\Desktop\bookml\books.csv')
users = pd.read_csv(r'C:\Users\vibha\OneDrive\Desktop\bookml\users.csv')
ratings = pd.read_csv(r'C:\Users\vibha\OneDrive\Desktop\bookml\ratings.csv
```

```
C:\Users\vibha\AppData\Local\Temp\ipykernel_25060\4165247160.py:1: DtypeW
arning: Columns (3) have mixed types. Specify dtype option on import or s
et low_memory=False.
  books = pd.read_csv(r'C:\Users\vibha\OneDrive\Desktop\bookml\books.cs
v')
```

In [4]:
```
print(books)
```

```
             ISBN                                          Book-Title  \
0       0195153448                                Classical Mythology
1       0002005018                                       Clara Callan
2       0060973129                              Decision in Normandy
3       0374157065  Flu: The Story of the Great Influenza Pandemic...
4       0393045218                            The Mummies of Urumchi
...            ...                                               ...
271355  0440400988                       There's a Bat in Bunk Five
271356  0525447644                          From One to One Hundred
271357  006008667X  Lily Dale : The True Story of the Town that Ta...
271358  0192126040                       Republic (World's Classics)
271359  0767409752  A Guided Tour of Rene Descartes' Meditations o...

                Book-Author Year-Of-Publication  \
0         Mark P. O. Morford                2002
1       Richard Bruce Wright                2001
2               Carlo D'Este                1991
3          Gina Bari Kolata                 1999
4           E. J. W. Barber                 1999
...                     ...                 ...
271355        Paula Danziger                1988
271356            Teri Sloat                1991
271357      Christine Wicker                2004
271358                 Plato                1996
271359   Christopher  Biffle                2000

                                        Publisher  \
0                          Oxford University Press
1                            HarperFlamingo Canada
2                                  HarperPerennial
3                            Farrar Straus Giroux
4                       W. W. Norton &amp;amp; Company
...                                           ...
271355                 Random House Childrens Pub (Mm)
271356                                 Dutton Books
271357                            HarperSanFrancisco
271358                         Oxford University Press
271359  McGraw-Hill Humanities/Social Sciences/Languages

                                             Image-URL-S  \
0       http://images.amazon.com/images/P/0195153448.0... (http://images.
amazon.com/images/P/0195153448.0...)
1       http://images.amazon.com/images/P/0002005018.0... (http://images.
amazon.com/images/P/0002005018.0...)
2       http://images.amazon.com/images/P/0060973129.0... (http://images.
amazon.com/images/P/0060973129.0...)
3       http://images.amazon.com/images/P/0374157065.0... (http://images.
amazon.com/images/P/0374157065.0...)
4       http://images.amazon.com/images/P/0393045218.0... (http://images.
amazon.com/images/P/0393045218.0...)
...                                                   ...
271355  http://images.amazon.com/images/P/0440400988.0... (http://images.
amazon.com/images/P/0440400988.0...)
271356  http://images.amazon.com/images/P/0525447644.0... (http://images.
amazon.com/images/P/0525447644.0...)
271357  http://images.amazon.com/images/P/006008667X.0... (http://images.
amazon.com/images/P/006008667X.0...)
271358  http://images.amazon.com/images/P/0192126040.0... (http://images.
amazon.com/images/P/0192126040.0...)
271359  http://images.amazon.com/images/P/0767409752.0... (http://images.
amazon.com/images/P/0767409752.0...)
```

```
                                                  Image-URL-M  \
0       http://images.amazon.com/images/P/0195153448.0... (http://images.
amazon.com/images/P/0195153448.0...)
1       http://images.amazon.com/images/P/0002005018.0... (http://images.
amazon.com/images/P/0002005018.0...)
2       http://images.amazon.com/images/P/0060973129.0... (http://images.
amazon.com/images/P/0060973129.0...)
3       http://images.amazon.com/images/P/0374157065.0... (http://images.
amazon.com/images/P/0374157065.0...)
4       http://images.amazon.com/images/P/0393045218.0... (http://images.
amazon.com/images/P/0393045218.0...)
...                                                   ...
271355  http://images.amazon.com/images/P/0440400988.0... (http://images.
amazon.com/images/P/0440400988.0...)
271356  http://images.amazon.com/images/P/0525447644.0... (http://images.
amazon.com/images/P/0525447644.0...)
271357  http://images.amazon.com/images/P/006008667X.0... (http://images.
amazon.com/images/P/006008667X.0...)
271358  http://images.amazon.com/images/P/0192126040.0... (http://images.
amazon.com/images/P/0192126040.0...)
271359  http://images.amazon.com/images/P/0767409752.0... (http://images.
amazon.com/images/P/0767409752.0...)

                                                  Image-URL-L
0       http://images.amazon.com/images/P/0195153448.0... (http://images.
amazon.com/images/P/0195153448.0...)
1       http://images.amazon.com/images/P/0002005018.0... (http://images.
amazon.com/images/P/0002005018.0...)
2       http://images.amazon.com/images/P/0060973129.0... (http://images.
amazon.com/images/P/0060973129.0...)
3       http://images.amazon.com/images/P/0374157065.0... (http://images.
amazon.com/images/P/0374157065.0...)
4       http://images.amazon.com/images/P/0393045218.0... (http://images.
amazon.com/images/P/0393045218.0...)
...                                                   ...
271355  http://images.amazon.com/images/P/0440400988.0... (http://images.
amazon.com/images/P/0440400988.0...)
271356  http://images.amazon.com/images/P/0525447644.0... (http://images.
amazon.com/images/P/0525447644.0...)
271357  http://images.amazon.com/images/P/006008667X.0... (http://images.
amazon.com/images/P/006008667X.0...)
271358  http://images.amazon.com/images/P/0192126040.0... (http://images.
amazon.com/images/P/0192126040.0...)
271359  http://images.amazon.com/images/P/0767409752.0... (http://images.
amazon.com/images/P/0767409752.0...)

[271360 rows x 8 columns]
```

In [5]:
```
print(users)
```

```
        User-ID                            Location   Age
0             1                  nyc, new york, usa   NaN
1             2            stockton, california, usa  18.0
2             3      moscow, yukon territory, russia   NaN
3             4            porto, v.n.gaia, portugal  17.0
4             5   farnborough, hants, united kingdom   NaN
...         ...                                  ...   ...
278853   278854               portland, oregon, usa   NaN
278854   278855  tacoma, washington, united kingdom  50.0
278855   278856            brampton, ontario, canada   NaN
278856   278857            knoxville, tennessee, usa   NaN
278857   278858                 dublin, n/a, ireland   NaN

[278858 rows x 3 columns]
```

In [6]:
```
print(ratings
     )
```

```
         User-ID         ISBN  Book-Rating
0         276725   034545104X            0
1         276726   0155061224            5
2         276727   0446520802            0
3         276729   052165615X            3
4         276729   0521795028            6
...          ...          ...          ...
1149775   276704   1563526298            9
1149776   276706   0679447156            0
1149777   276709   0515107662           10
1149778   276721   0590442449           10
1149779   276723  05162443314            8

[1149780 rows x 3 columns]
```

In [3]:
```
books['Image-URL-M'][1]
```

Out[3]: 'http://images.amazon.com/images/P/0002005018.01.MZZZZZZZ.jpg'

In [4]:
```
users.head()
```

Out[4]:

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

In [5]:
```
print(books.shape)
print(ratings.shape)
print(users.shape)
```

```
(271360, 8)
(1149780, 3)
(278858, 3)
```

In [6]:
```
books.isnull().sum()
```

Out[6]:
```
ISBN                 0
Book-Title           0
Book-Author          1
Year-Of-Publication  0
Publisher            2
Image-URL-S          0
Image-URL-M          0
Image-URL-L          3
dtype: int64
```

In [7]:
```
users.isnull().sum()
```

Out[7]:
```
User-ID         0
Location        0
Age        110762
dtype: int64
```

In [8]:
```
ratings.isnull().sum()
```

Out[8]:
```
User-ID        0
ISBN           0
Book-Rating    0
dtype: int64
```

In [9]:
```
books.duplicated().sum()
```

Out[9]:
```
0
```

In [10]:
```
ratings.duplicated().sum()
```

Out[10]:
```
0
```

In [11]:
```
users.duplicated().sum()
```

Out[11]:
```
0
```

In [12]:
```
ratings_with_name = ratings.merge(books,on='ISBN')
```

In [13]:
```
num_rating_df = ratings_with_name.groupby('Book-Title').count()['Book-Rati
num_rating_df.rename(columns={'Book-Rating':'num_ratings'},inplace=True)
num_rating_df
```

Out[13]:

	Book-Title	num_ratings
0	A Light in the Storm: The Civil War Diary of ...	4
1	Always Have Popsicles	1
2	Apple Magic (The Collector's series)	1
3	Ask Lily (Young Women of Faith: Lily Series, ...	1
4	Beyond IBM: Leadership Marketing and Finance ...	1
...	...	...
241066	Ã?Â?lpiraten.	2
241067	Ã?Â?rger mit Produkt X. Roman.	4
241068	Ã?Â?sterlich leben.	1
241069	Ã?Â?stlich der Berge.	3
241070	Ã?Â?thique en toc	2

241071 rows × 2 columns

In [14]:
```
avg_rating_df = ratings_with_name.groupby('Book-Title').mean()['Book-Ratin
avg_rating_df.rename(columns={'Book-Rating':'avg_rating'},inplace=True)
avg_rating_df
```

Out[14]:

	Book-Title	avg_rating
0	A Light in the Storm: The Civil War Diary of ...	2.250000
1	Always Have Popsicles	0.000000
2	Apple Magic (The Collector's series)	0.000000
3	Ask Lily (Young Women of Faith: Lily Series, ...	8.000000
4	Beyond IBM: Leadership Marketing and Finance ...	0.000000
...	...	...
241066	Ã?Â?lpiraten.	0.000000
241067	Ã?Â?rger mit Produkt X. Roman.	5.250000
241068	Ã?Â?sterlich leben.	7.000000
241069	Ã?Â?stlich der Berge.	2.666667
241070	Ã?Â?thique en toc	4.000000

241071 rows × 2 columns

In [15]:
```
popular_df = num_rating_df.merge(avg_rating_df,on='Book-Title')
popular_df
```

Out[15]:

	Book-Title	num_ratings	avg_rating
0	A Light in the Storm: The Civil War Diary of ...	4	2.250000
1	Always Have Popsicles	1	0.000000
2	Apple Magic (The Collector's series)	1	0.000000
3	Ask Lily (Young Women of Faith: Lily Series, ...	1	8.000000
4	Beyond IBM: Leadership Marketing and Finance ...	1	0.000000
...	...	...	...
241066	Ã?Â?lpiraten.	2	0.000000
241067	Ã?Â?rger mit Produkt X. Roman.	4	5.250000
241068	Ã?Â?sterlich leben.	1	7.000000
241069	Ã?Â?stlich der Berge.	3	2.666667
241070	Ã?Â?thique en toc	2	4.000000

241071 rows × 3 columns

In [16]:
```
popular_df = popular_df[popular_df['num_ratings']>=250].sort_values('avg_r
```

In [17]:
```
popular_df = popular_df.merge(books,on='Book-Title').drop_duplicates('Book
```

In [18]:
```
popular_df['Image-URL-M'][0]
```

Out[18]:
```
'http://images.amazon.com/images/P/0439136350.01.MZZZZZZZ.jpg'
```

In [19]:
```
x = ratings_with_name.groupby('User-ID').count()['Book-Rating'] > 200
padhe_likhe_users = x[x].index
```

In [20]:
```
filtered_rating = ratings_with_name[ratings_with_name['User-ID'].isin(padh
```

In [21]:
```
y = filtered_rating.groupby('Book-Title').count()['Book-Rating']>=50
famous_books = y[y].index
```

In [22]:
```
final_ratings = filtered_rating[filtered_rating['Book-Title'].isin(famous_
```

In [23]:
```
pt = final_ratings.pivot_table(index='Book-Title',columns='User-ID',values
```

In [24]:
```
pt.fillna(0,inplace=True)
```

In [25]:
```
pt
```

Out[25]:

User-ID	254	2276	2766	2977	3363	4017	4385	6251	6323	6543	...	271705	273
Book-Title													
1984	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	10.0	
1st to Die: A Novel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	...	0.0	
2nd Chance	0.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
4 Blondes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
A Bend in the Road	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
Year of Wonders	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
You Belong To Me	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
Zen and the Art of Motorcycle Maintenance: An Inquiry into Values	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
Zoya	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
\O\" Is for Outlaw"	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	

706 rows × 810 columns

In [26]:
```
from sklearn.metrics.pairwise import cosine_similarity
```

In [27]:
```
similarity_scores = cosine_similarity(pt)
```

In [28]:
```
similarity_scores.shape
```

Out[28]:
```
(706, 706)
```

In [29]:
```
def recommend(book_name):
    # index fetch
    index = np.where(pt.index==book_name)[0][0]
    similar_items = sorted(list(enumerate(similarity_scores[index])),key=l

    data = []
    for i in similar_items:
        item = []
        temp_df = books[books['Book-Title'] == pt.index[i[0]]]
        item.extend(list(temp_df.drop_duplicates('Book-Title')['Book-Title
        item.extend(list(temp_df.drop_duplicates('Book-Title')['Book-Autho
        item.extend(list(temp_df.drop_duplicates('Book-Title')['Image-URL-

        data.append(item)

    return data
```

In [30]:
```
recommend('1984')
```

Out[30]:
```
[['Animal Farm',
  'George Orwell',
  'http://images.amazon.com/images/P/0451526341.01.MZZZZZZZ.jpg'],
 ["The Handmaid's Tale",
  'Margaret Atwood',
  'http://images.amazon.com/images/P/0449212602.01.MZZZZZZZ.jpg'],
 ['Brave New World',
  'Aldous Huxley',
  'http://images.amazon.com/images/P/0060809833.01.MZZZZZZZ.jpg'],
 ['The Vampire Lestat (Vampire Chronicles, Book II)',
  'ANNE RICE',
  'http://images.amazon.com/images/P/0345313860.01.MZZZZZZZ.jpg']]
```

In [31]:
```
pt.index[545]
```

Out[31]:
```
"The Handmaid's Tale"
```

In [32]:
```
import pickle
pickle.dump(popular_df,open('popular.pkl','wb'))
```

In [33]:
```
books.drop_duplicates('Book-Title')
```

Out[33]:

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.a
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.a
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.a
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.a
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.a
...	...	...	...	...	...	
271354	0449906736	Flashpoints: Promise and Peril in a New World	Robin Wright	1993	Ballantine Books	http://images.a
271356	0525447644	From One to One Hundred	Teri Sloat	1991	Dutton Books	http://images.a
271357	006008667X	Lily Dale : The True Story of the Town that Ta...	Christine Wicker	2004	HarperSanFrancisco	http://images.a
271358	0192126040	Republic (World's Classics)	Plato	1996	Oxford University Press	http://images.a
271359	0767409752	A Guided Tour of Rene Descartes' Meditations o...	Christopher Biffle	2000	McGraw-Hill Humanities/Social Sciences/Languages	http://images.a

242135 rows × 8 columns

In [34]:
```
pickle.dump(pt,open('pt.pkl','wb'))
pickle.dump(books,open('books.pkl','wb'))
pickle.dump(similarity_scores,open('similarity_scores.pkl','wb'))
```

In [ ]:
```

```