```
In [1]:
           import numpy as np
           import pandas as pd
In [2]:
           books=pd.read_csv('Books.csv')
           users=pd.read_csv('Users.csv')
           ratings=pd.read_csv('Ratings.csv')
          C:\Users\KIRANS~1\AppData\Local\Temp/ipykernel_4556/3397335393.py:1: DtypeWarning: Columns (3) have mixed types.
          Specify dtype option on import or set low_memory=False.
            books=pd.read_csv('Books.csv')
In [3]:
           books.head()
Out[3]:
                                      Book-
                                               Year-Of-
                   ISBN
                         Book-Title
                                                              Publisher
                                                                                                        Image-URL-S
                                     Author
                                            Publication
                                    Mark P.
                           Classical
                                                                Oxford
          0 0195153448
                                         0
                                                   2002
                                                                        http://images.amazon.com/images/P/0195153448.0... http://images.amazon.com/images/l
                                                        University Press
                          Mythology
                                     Morford
                                     Richard
                              Clara
                                                         HarperFlamingo
          1 0002005018
                                                   2001
                                                                       http://images.amazon.com/images/P/0002005018.0... http://images.amazon.com/images/l
                                      Bruce
                             Callan
                                                               Canada
                                      Wright
                          Decision in
                                      Carlo
          2 0060973129
                                                   1991 HarperPerennial http://images.amazon.com/images/P/0060973129.0... http://images.amazon.com/images/l
                                     D'Este
                          Normandy
                            Flu: The
                                       Gina
                         Story of the
                                                           Farrar Straus
          3 0374157065
                                                   1999
                                                                       http://images.amazon.com/images/P/0374157065.0... http://images.amazon.com/images/l
                              Great
                                       Bari
                                                                Giroux
                           Influenza
                                      Kolata
                         Pandemic...
                               The
                                       E. J.
                                                           W. W. Norton
          4 0393045218
                                                   1999
                          Mummies
                                         W.
                                                                       http://images.amazon.com/images/P/0393045218.0... http://images.amazon.com/images/l
                                                                &
                          of Urumchi
                                     Barber
                                                              Company
In [4]:
           users.head()
Out[4]:
             User-ID
                                           Location Age
                                   nyc, new york, usa NaN
                  2
                               stockton, california, usa
          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]:
           ratings.head()
                           ISBN Book-Rating
Out[5]:
             User-ID
            276725 034545104X
                                           0
             276726 0155061224
                                           5
             276727 0446520802
                                           0
                                           3
             276729 052165615X
                                           6
          4 276729 0521795028
In [6]:
           print(books.shape)
           print(ratings.shape)
           print(users.shape)
          (271360, 8)
          (1149780, 3)
          (278858, 3)
```

In [7]: books.isnull().sum()

```
ISBN
 Out[7]:
          Book-Title
                                     0
           Book-Author
                                     1
           Year-Of-Publication
           Publisher
                                     2
           {\tt Image-URL-S}
                                     0
           Image-URL-M
           Image-URL-L
                                     3
          dtype: int64
 In [8]:
           users.isnull().sum()
          User-ID
                              0
 Out[8]:
          Location
                              0
          Age
                        110762
          dtype: int64
 In [9]:
             ratings.isnull().sum()
          User-ID
 Out[9]:
          ISBN
                            0
           Book-Rating
                            0
          dtype: int64
In [10]:
            books.duplicated().sum()
Out[10]:
In [11]:
           users.duplicated().sum()
Out[11]:
In [12]:
            ratings.duplicated().sum()
Out[12]:
In [13]:
            rating_with_name=ratings.merge(books,on='ISBN')
In [14]:
           num rating df=rating with name.groupby('Book-Title').count()['Book-Rating'].reset index()
           num_rating_df.rename(columns={'Book-Rating':'rating'},inplace=True)
           num_rating_df
Out[14]:
                                                 Book-Title rating
               0
                      A Light in the Storm: The Civil War Diary of ...
                                       Always Have Popsicles
               2
                             Apple Magic (The Collector's series)
               3
                     Ask Lily (Young Women of Faith: Lily Series, ...
                4 Beyond IBM: Leadership Marketing and Finance ...
                                                               2
           241066
                                               Ã?Â?lpiraten.
           241067
                                Ã?Â?rger mit Produkt X. Roman.
           241068
                                           Ã?Â?sterlich leben.
                                                                1
           241069
                                                               3
                                         Ã?Â?stlich der Berge.
           241070
                                            Ã?Â?thique en toc
          241071 rows × 2 columns
```

```
avg rating df=rating with name.groupby('Book-Title').mean()['Book-Rating'].reset index()
            avg_rating_df.rename(columns={'Book-Rating':'average-rating'},inplace=True)
            avg rating df
                                                     Book-Title average-rating
Out[15]:
                 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
                                                   Ã?Â?lpiraten.
            241066
                                                                      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 [16]:
            popular df = num rating df.merge(avg rating df,on='Book-Title')
            popular_df
                                                     Book-Title rating average-rating
Out[16]:
                                                                             2.250000
                 0
                        A Light in the Storm: The Civil War Diary of ...
                 1
                                           Always Have Popsicles
                                                                             0.000000
                                Apple Magic (The Collector's series)
                 2
                                                                             0.000000
                                                                     1
                       Ask Lily (Young Women of Faith: Lily Series, ...
                 3
                                                                             8.000000
                 4 Beyond IBM: Leadership Marketing and Finance ...
                                                                             0.000000
                                                                     1
            241066
                                                   Ã?Â?lpiraten.
                                                                     2
                                                                             0.000000
            241067
                                   Ã?Â?rger mit Produkt X. Roman.
                                                                             5.250000
                                              Ã?Â?sterlich leben.
            241068
                                                                     1
                                                                             7 000000
            241069
                                            Ã?Â?stlich der Berge.
                                                                     3
                                                                             2.666667
            241070
                                               Ã?Â?thique en toc
                                                                             4.000000
                                                                     2
          241071 rows × 3 columns
In [17]:
            popular df=popular df[popular df['rating']>=200].sort values('average-rating',ascending=False).head(30)
```

```
In [18]:
```

In [19]: popular df

Out[19]:

:	Book-Title	Book-Author	Image-URL-M	rating	average- rating
0	Harry Potter and the Prisoner of Azkaban (Book 3)	J. K. Rowling	http://images.amazon.com/images/P/0439136350.0	428	5.852804
3	Harry Potter and the Goblet of Fire (Book 4)	J. K. Rowling	http://images.amazon.com/images/P/0439139597.0	387	5.824289
5	Harry Potter and the Sorcerer's Stone (Book 1)	J. K. Rowling	http://images.amazon.com/images/P/0590353403.0	278	5.737410
9	Harry Potter and the Order of the Phoenix (Boo	J. K. Rowling	http://images.amazon.com/images/P/043935806X.0	347	5.501441
13	Ender's Game (Ender Wiggins Saga (Paperback))	Orson Scott Card	http://images.amazon.com/images/P/0312853238.0	249	5.409639
17	Harry Potter and the Chamber of Secrets (Book 2)	J. K. Rowling	http://images.amazon.com/images/P/0439064872.0	556	5.183453
20	The Hobbit : The Enchanting Prelude to The Lor	J.R.R. TOLKIEN	http://images.amazon.com/images/P/0345339681.0	281	5.007117
21	The Fellowship of the Ring (The Lord of the Ri	J.R.R. TOLKIEN	http://images.amazon.com/images/P/0345339703.0	368	4.948370
30	Harry Potter and the Sorcerer's Stone (Harry P	J. K. Rowling	http://images.amazon.com/images/P/059035342X.0	575	4.895652
32	The Two Towers (The Lord of the Rings, Part 2)	J.R.R. TOLKIEN	http://images.amazon.com/images/P/0345339711.0	260	4.880769
43	To Kill a Mockingbird	Harper Lee	http://images.amazon.com/images/P/0446310786.0	510	4.700000

51	The Da Vinci Code	Dan Brown	http://images.amazon.com/images/P/0385504209.0	898	4.642539
57	The Five People You Meet in Heaven	Mitch Albom	http://images.amazon.com/images/P/0786868716.0	430	4.551163
59	The Catcher in the Rye	J.D. Salinger	http://images.amazon.com/images/P/0316769487.0	449	4.545657
66	The Return of the King (The Lord of the Rings,	J.R.R. TOLKIEN	http://images.amazon.com/images/P/0345339738.0	211	4.497630
76	The Lovely Bones: A Novel	Alice Sebold	http://images.amazon.com/images/P/0316666343.0	1295	4.468726
77	The Great Gatsby	F. Scott Fitzgerald	http://images.amazon.com/images/P/185326041X.0	205	4.458537
79	1984	George Orwell	http://images.amazon.com/images/P/0451524934.0	284	4.454225
88	Prodigal Summer: A Novel	Barbara Kingsolver	http://images.amazon.com/images/P/0060959037.0	253	4.450593
89	Neverwhere	Neil Gaiman	http://images.amazon.com/images/P/0380789019.0	265	4.449057
94	The Secret Life of Bees	Sue Monk Kidd	http://images.amazon.com/images/P/0142001740.0	774	4.447028
100	Stupid White Menand Other Sorry Excuses fo	Michael Moore	http://images.amazon.com/images/P/0060392452.0	283	4.356890
101	Tuesdays with Morrie: An Old Man, a Young Man,	MITCH ALBOM	http://images.amazon.com/images/P/0385484518.0	493	4.354970
104	The Red Tent (Bestselling Backlist)	Anita Diamant	http://images.amazon.com/images/P/0312195516.0	723	4.334716
105	Brave New World	Aldous Huxley	http://images.amazon.com/images/P/0060809833.0	226	4.331858
110	The Hitchhiker's Guide to the Galaxy	Douglas Adams	http://images.amazon.com/images/P/0671461494.0	268	4.328358
119	Nickel and Dimed: On (Not) Getting By in America	Barbara Ehrenreich	http://images.amazon.com/images/P/0805063897.0	335	4.289552
121	The Tao of Pooh	Benjamin Hoff	http://images.amazon.com/images/P/0140067477.0	211	4.289100
123	Animal Farm	George Orwell	http://images.amazon.com/images/P/0451526341.0	233	4.274678
133	Into the Wild	Jon Krakauer	http://images.amazon.com/images/P/0385486804.0	252	4.273810

Collaborative Filtering

```
In [20]:
           variable=rating_with_name.groupby('User-ID').count()['Book-Rating'] >150
           padha_likhe_users=variable[variable].index
In [21]:
            filtered rating=rating with name[rating with name['User-ID'].isin(padha likhe users)]
In [22]:
           variable2=filtered rating.groupby('Book-Title').count()['Book-Rating']>=30
            famous books=variable2[variable2].index
In [23]:
           final_ratings=filtered_rating[filtered_rating['Book-Title'].isin(famous_books)]
In [24]:
           #final_ratings.drop_duplicates()
           pt = final_ratings.pivot_table(index='Book-Title',columns='User-ID',values='Book-Rating')
In [25]:
           pt.fillna(0,inplace=True)
In [26]:
                User-ID 254 1733 1903 2033 2110 2276 2766 2891 2977 3363 ... 274301 274308 275970 276680 277427 277478 27763
Out[26]:
             Book-Title
                 10 Lb.
                                                                                                                                            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
                                                                                                                                     0.0
                Penalty
                    16
            Lighthouse
                        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
                                                                                                                                            0
                  Road
                  1984
                        9.0
                                                                           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
            1st to Die: A
                              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
                                                                      0.0
                 Novel
                 2010:
                                                                           0.0 ...
               Odyssey
                        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
            You Belong
                        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
                                                                                                                                            0
                 To Me
            Zen and the
```

```
Art of
            Motorcycle
                       0.0
                            0.0
                                                        0.0
                                                                        0.0 ...
                                                                                          0.0
          Maintenance:
             An Inquiry
            into Values
          Zlata's Diary:
          A Child's Life
                       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
                                                                                                                                      0
            in Sarajevo
                 Zoya
                       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
                                                                                                                                      0
             \O\" Is for
                       0.0
                            0.0
                                  0.0
                                       0.0
                                             0.0
                                                  0.0
                                                        0.0
                                                             0.0
                                                                  0.0
                                                                        0.0 ...
                                                                                  8.0
                                                                                          0.0
                                                                                                 0.0
                                                                                                        0.0
                                                                                                                0.0
                                                                                                                       0.0
                                                                                                                              0.0
                                                                                                                                      0
         1983 rows × 1113 columns
In [27]:
           from sklearn.metrics.pairwise import cosine_similarity
In [28]:
           similarity score=cosine similarity(pt)
In [29]:
           similarity_score.shape
          (1983, 1983)
Out[29]:
In [30]:
           def recommend(book name):
               # index fetch
               index = np.where(pt.index==book_name)[0][0]
               similar_items = sorted(list(enumerate(similarity_score[index])),key=lambda x:x[1],reverse=True)[1:5]
               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'].values))
                    item.extend(list(temp_df.drop_duplicates('Book-Title')['Book-Author'].values))
                    item.extend(list(temp df.drop duplicates('Book-Title')['Image-URL-M'].values))
                    data.append(item)
               return data
In [31]:
           #recommend('Message in a Bottle')
           #recommend('The Notebook')
           #recommend('1984')
           recommend('Animal Farm')
Out[31]: [['Hamlet',
             'William Shakespeare',
             'http://images.amazon.com/images/P/067172262X.01.MZZZZZZZ.jpg'],
           ['Women in His Life'
             'Barbara Taylor Bradford',
             'http://images.amazon.com/images/P/0345345738.01.MZZZZZZZ.jpg'],
           ['Before and After',
             'Rosellen Brown',
             'http://images.amazon.com/images/P/0440216540.01.MZZZZZZZ.jpg'],
           ['Poland',
             'James A. Michener'.
             'http://images.amazon.com/images/P/0394533887.01.MZZZZZZZ.jpg']]
In [32]:
           pt.index[545]
          'Flowers for Algernon (Bantam Classic)'
Out[32]:
In [33]:
           import pickle
           pickle.dump(popular_df,open('popular.pkl','wb'))
In [34]:
           books.drop_duplicates('Book-Title')
Out[34]:
                                           Book-
                                                    Year-Of-
                                                                                                            Image-URL-S
                       ISBN Book-Title
                                                                     Publisher
                                           Author Publication
```

0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0	http://images.amazo
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0	http://images.amazo
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0	http://images.amazo
3	0374157065	Flu: The Story of the Great Influenza Pandemic	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0	http://images.amazo
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0	http://images.amazc
271354	0449906736	Flashpoints: Promise and Peril in a New World	Robin Wright	1993	Ballantine Books	http://images.amazon.com/images/P/0449906736.0	http://images.amazo
271356	0525447644	From One to One Hundred	Teri Sloat	1991	Dutton Books	http://images.amazon.com/images/P/0525447644.0	http://images.amazo
271357	006008667X	Lily Dale : The True Story of the Town that Ta	Christine Wicker	2004	HarperSanFrancisco	http://images.amazon.com/images/P/006008667X.0	http://images.amazc
271358	0192126040	Republic (World's Classics)	Plato	1996	Oxford University Press	http://images.amazon.com/images/P/0192126040.0	http://images.amazo
271359	0767409752	A Guided Tour of Rene Descartes' Meditations 0	Christopher Biffle	2000	McGraw-Hill Humanities/Social Sciences/Languages	http://images.amazon.com/images/P/0767409752.0	http://images.amazc
242135 rows × 8 columns							

```
in [35]:
    pickle.dump(pt,open('pt.pkl','wb'))
    pickle.dump(books,open('books.pkl','wb'))
    pickle.dump(similarity_score,open('similarity_score.pkl','wb'))
```

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js