```
import numpy as np
import pandas as pd
```

We perform some Exploratory Data Analysis.

```
cd C:\Users\soumy\Downloads\archive (6)
```

```
C:\Users\soumy\Downloads\archive (6)
```

```
books = pd.read_csv('books.csv')
users = pd.read_csv('users.csv')
ratings = pd.read_csv('ratings.csv')
```

```
C:\Users\soumy\AppData\Local\Temp\ipykernel_17172\3819754567.py:1:
DtypeWarning: Columns (3) have mixed types. Specify dtype option on
import or set low_memory=False.
  books = pd.read_csv('books.csv')
```

```
books['Image-URL-M'][1]
```

```
'http://images.amazon.com/images/P/0002005018.01.MZZZZZZZ.jpg'
```

```
books.head(10)
```

```
         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
5  0399135782                             The Kitchen God's Wife
6  0425176428  What If?: The World's Foremost Military Histor...
7  0671870432                                    PLEADING GUILTY
8  0679425608  Under the Black Flag: The Romance and the Real...
9  074322678X            Where You'll Find Me: And Other Stories

            Book-Author Year-Of-Publication
Publisher
0    Mark P. O. Morford                2002     Oxford University
Press  \
1  Richard Bruce Wright                2001       HarperFlamingo
Canada
2          Carlo D'Este                1991
HarperPerennial
3      Gina Bari Kolata                1999        Farrar Straus
Giroux
4       E. J. W. Barber                1999  W. W. Norton &amp;
Company
5               Amy Tan                1991            Putnam Pub
Group
6         Robert Cowley                2000    Berkley Publishing
```

```
Group
7           Scott Turow                1993
Audioworks
8       David Cordingly                1996                Random
House
9           Ann Beattie                2002
Scribner

                                         Image-URL-S
0  http://images.amazon.com/images/P/0195153448.0...  \
1  http://images.amazon.com/images/P/0002005018.0...
2  http://images.amazon.com/images/P/0060973129.0...
3  http://images.amazon.com/images/P/0374157065.0...
4  http://images.amazon.com/images/P/0393045218.0...
5  http://images.amazon.com/images/P/0399135782.0...
6  http://images.amazon.com/images/P/0425176428.0...
7  http://images.amazon.com/images/P/0671870432.0...
8  http://images.amazon.com/images/P/0679425608.0...
9  http://images.amazon.com/images/P/074322678X.0...

                                         Image-URL-M
0  http://images.amazon.com/images/P/0195153448.0...  \
1  http://images.amazon.com/images/P/0002005018.0...
2  http://images.amazon.com/images/P/0060973129.0...
3  http://images.amazon.com/images/P/0374157065.0...
4  http://images.amazon.com/images/P/0393045218.0...
5  http://images.amazon.com/images/P/0399135782.0...
6  http://images.amazon.com/images/P/0425176428.0...
7  http://images.amazon.com/images/P/0671870432.0...
8  http://images.amazon.com/images/P/0679425608.0...
9  http://images.amazon.com/images/P/074322678X.0...

                                         Image-URL-L
0  http://images.amazon.com/images/P/0195153448.0...
1  http://images.amazon.com/images/P/0002005018.0...
2  http://images.amazon.com/images/P/0060973129.0...
3  http://images.amazon.com/images/P/0374157065.0...
4  http://images.amazon.com/images/P/0393045218.0...
5  http://images.amazon.com/images/P/0399135782.0...
6  http://images.amazon.com/images/P/0425176428.0...
7  http://images.amazon.com/images/P/0671870432.0...
8  http://images.amazon.com/images/P/0679425608.0...
9  http://images.amazon.com/images/P/074322678X.0...
```

```
ratings.head(10)
```

```
   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
5   276733  2080674722            0
6   276736  3257224281            8
7   276737  0600570967            6
8   276744  038550120X            7
9   276745   342310538           10
```

```
users.head(10)
```

```
   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
5        6        santa monica, california, usa  61.0
6        7                  washington, dc, usa   NaN
7        8             timmins, ontario, canada   NaN
8        9           germantown, tennessee, usa   NaN
9       10           albacete, wisconsin, spain  26.0
```

```
print(books.shape)
print(ratings.shape)
print(users.shape)
```

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

```
books.isnull().sum()
```

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

```
users.isnull().sum()
```

```
User-ID          0
Location         0
Age         110762
dtype: int64
```

```
ratings.isnull().sum()
```

```
User-ID        0
ISBN           0
Book-Rating    0
dtype: int64
```

```
books.duplicated().sum()
```

```
0
```

```
users.duplicated().sum()
```

```
0
```

```
ratings.duplicated().sum()
```

```
0
```

```
ratings_with_name = ratings.merge(books,on='ISBN')
```

```
ratings_with_name.head(10)
```

```
   User-ID        ISBN  Book-Rating            Book-Title Book-Author

0   276725  034545104X            0  Flesh Tones: A Novel  M. J. Rose
\
1     2313  034545104X            5  Flesh Tones: A Novel  M. J. Rose

2     6543  034545104X            0  Flesh Tones: A Novel  M. J. Rose

3     8680  034545104X            5  Flesh Tones: A Novel  M. J. Rose

4    10314  034545104X            9  Flesh Tones: A Novel  M. J. Rose

5    23768  034545104X            0  Flesh Tones: A Novel  M. J. Rose

6    28266  034545104X            0  Flesh Tones: A Novel  M. J. Rose

7    28523  034545104X            0  Flesh Tones: A Novel  M. J. Rose

8    39002  034545104X            0  Flesh Tones: A Novel  M. J. Rose

9    50403  034545104X            9  Flesh Tones: A Novel  M. J. Rose


  Year-Of-Publication         Publisher
0                 2002  Ballantine Books  \
1                 2002  Ballantine Books
2                 2002  Ballantine Books
3                 2002  Ballantine Books
4                 2002  Ballantine Books
5                 2002  Ballantine Books
6                 2002  Ballantine Books
```

```
7                2002  Ballantine Books
8                2002  Ballantine Books
9                2002  Ballantine Books

                                         Image-URL-S
0  http://images.amazon.com/images/P/034545104X.0...  \
1  http://images.amazon.com/images/P/034545104X.0...
2  http://images.amazon.com/images/P/034545104X.0...
3  http://images.amazon.com/images/P/034545104X.0...
4  http://images.amazon.com/images/P/034545104X.0...
5  http://images.amazon.com/images/P/034545104X.0...
6  http://images.amazon.com/images/P/034545104X.0...
7  http://images.amazon.com/images/P/034545104X.0...
8  http://images.amazon.com/images/P/034545104X.0...
9  http://images.amazon.com/images/P/034545104X.0...

                                         Image-URL-M
0  http://images.amazon.com/images/P/034545104X.0...  \
1  http://images.amazon.com/images/P/034545104X.0...
2  http://images.amazon.com/images/P/034545104X.0...
3  http://images.amazon.com/images/P/034545104X.0...
4  http://images.amazon.com/images/P/034545104X.0...
5  http://images.amazon.com/images/P/034545104X.0...
6  http://images.amazon.com/images/P/034545104X.0...
7  http://images.amazon.com/images/P/034545104X.0...
8  http://images.amazon.com/images/P/034545104X.0...
9  http://images.amazon.com/images/P/034545104X.0...

                                         Image-URL-L
0  http://images.amazon.com/images/P/034545104X.0...
1  http://images.amazon.com/images/P/034545104X.0...
2  http://images.amazon.com/images/P/034545104X.0...
3  http://images.amazon.com/images/P/034545104X.0...
4  http://images.amazon.com/images/P/034545104X.0...
5  http://images.amazon.com/images/P/034545104X.0...
6  http://images.amazon.com/images/P/034545104X.0...
7  http://images.amazon.com/images/P/034545104X.0...
8  http://images.amazon.com/images/P/034545104X.0...
9  http://images.amazon.com/images/P/034545104X.0...
```

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

```
                                               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 x 2 columns]
```

Now we built a Collaborative Filtering Based Recommender System.

In Collaborative Filtering, we tend to find similar users and recommend what similar users like. In this type of recommendation system, we don't ## use the features of the item to recommend it, rather we classify the users into clusters of similar types and recommend each user according to the ## preference of its cluster.

Filtering out only those users who have rated at least 200 books(avid readers, that is) and books which have more that 50 ratings(popular books.)

```
x = ratings_with_name.groupby('User-ID').count()['Book-Rating'] > 200
Avid_readers = x[x].index
filtered_rating = ratings_with_name[ratings_with_name['User-ID'].isin(Avid_readers)]

y = filtered_rating.groupby('Book-Title').count()['Book-Rating']>=50
pop_books = y[y].index

final_ratings = filtered_rating[filtered_rating['Book-Title'].isin(pop_books)]

pt = final_ratings.pivot_table(index='Book-Title',columns='User-ID',values='Book-Rating')
pt.fillna(0,inplace=True)
pt
```

```
User-ID                                              254     2276
2766
Book-Title
```

```
1984                                                   9.0     0.0
0.0  \
1st to Die: A Novel                                    0.0     0.0
0.0
2nd Chance                                             0.0    10.0
0.0
4 Blondes                                              0.0     0.0
0.0
A Bend in the Road                                     0.0     0.0
7.0
...                                                    ...     ...
...
Year of Wonders                                        0.0     0.0
0.0
You Belong To Me                                       0.0     0.0
0.0
Zen and the Art of Motorcycle Maintenance: An I...     0.0     0.0
0.0
Zoya                                                   0.0     0.0
0.0
\O\" Is for Outlaw"                                    0.0     0.0
0.0

User-ID                                             2977    3363
4017
Book-Title

1984                                                   0.0     0.0
0.0  \
1st to Die: A Novel                                    0.0     0.0
0.0
2nd Chance                                             0.0     0.0
0.0
4 Blondes                                              0.0     0.0
0.0
A Bend in the Road                                     0.0     0.0
0.0
...                                                    ...     ...
...
Year of Wonders                                        7.0     0.0
0.0
You Belong To Me                                       0.0     0.0
0.0
Zen and the Art of Motorcycle Maintenance: An I...     0.0     0.0
0.0
Zoya                                                   0.0     0.0
0.0
\O\" Is for Outlaw"                                    0.0     0.0
0.0
```

```
User-ID                                             4385    6251
6323
Book-Title

1984                                                   0.0     0.0
0.0  \
1st to Die: A Novel                                    0.0     0.0
0.0
2nd Chance                                             0.0     0.0
0.0
4 Blondes                                              0.0     0.0
0.0
A Bend in the Road                                     0.0     0.0
0.0
...                                                    ...     ...
...
Year of Wonders                                        0.0     0.0
0.0
You Belong To Me                                       0.0     0.0
0.0
Zen and the Art of Motorcycle Maintenance: An I...     0.0     0.0
0.0
Zoya                                                   0.0     0.0
0.0
\O\" Is for Outlaw"                                    0.0     0.0
0.0

User-ID                                             6543    ...
271705
Book-Title                                                  ...

1984                                                   0.0  ...
10.0  \
1st to Die: A Novel                                    9.0  ...
0.0
2nd Chance                                             0.0  ...
0.0
4 Blondes                                              0.0  ...
0.0
A Bend in the Road                                     0.0  ...
0.0
...                                                    ...  ...     ..
.
Year of Wonders                                        0.0  ...
0.0
You Belong To Me                                       0.0  ...
0.0
Zen and the Art of Motorcycle Maintenance: An I...     0.0  ...
0.0
```

```
Zoya                                                   0.0  ...
0.0
\O\" Is for Outlaw"                                     0.0  ...
0.0

User-ID                                             273979  274004
274061
Book-Title

1984                                                   0.0     0.0
0.0  \
1st to Die: A Novel                                    0.0     0.0
0.0
2nd Chance                                             0.0     0.0
0.0
4 Blondes                                              0.0     0.0
0.0
A Bend in the Road                                     0.0     0.0
0.0
...                                                    ...     ...
...
Year of Wonders                                        9.0     0.0
0.0
You Belong To Me                                       0.0     0.0
0.0
Zen and the Art of Motorcycle Maintenance: An I...     0.0     0.0
0.0
Zoya                                                   0.0     0.0
0.0
\O\" Is for Outlaw"                                     0.0     0.0
0.0

User-ID                                             274301  274308
275970
Book-Title

1984                                                   0.0     0.0
0.0  \
1st to Die: A Novel                                    0.0     0.0
0.0
2nd Chance                                             0.0     0.0
0.0
4 Blondes                                              0.0     0.0
0.0
A Bend in the Road                                     0.0     0.0
0.0
...                                                    ...     ...
...
Year of Wonders                                        0.0     0.0
0.0
```

```
You Belong To Me                                       0.0     0.0
0.0
Zen and the Art of Motorcycle Maintenance: An I...     0.0     0.0
0.0
Zoya                                                   0.0     0.0
0.0
\O\" Is for Outlaw"                                    8.0     0.0
0.0

User-ID                                             277427  277639
278418
Book-Title

1984                                                   0.0     0.0
0.0
1st to Die: A Novel                                    0.0     0.0
0.0
2nd Chance                                             0.0     0.0
0.0
4 Blondes                                              0.0     0.0
0.0
A Bend in the Road                                     0.0     0.0
0.0
...                                                    ...     ...
...
Year of Wonders                                        0.0     0.0
0.0
You Belong To Me                                       0.0     0.0
0.0
Zen and the Art of Motorcycle Maintenance: An I...     0.0     0.0
0.0
Zoya                                                   0.0     0.0
0.0
\O\" Is for Outlaw"                                    0.0     0.0
0.0

[706 rows x 810 columns]
```

To each book we can assign a vector of length 810, where the ith entry corresponds to the rating given to this book by user i.

We then canculate the mutual distance(Euclidean) between each pair of books and upon call, recommend those 5 books which are nearest.

```
from sklearn.metrics.pairwise import cosine_similarity


similarity_scores = cosine_similarity(pt)

similarity_scores

array([[1.        , 0.10255025, 0.01220856, ..., 0.12110367,
0.07347567,
        0.04316046],
       [0.10255025, 1.        , 0.2364573 , ..., 0.07446129,
0.16773875,
        0.14263397],
       [0.01220856, 0.2364573 , 1.        , ..., 0.04558758,
0.04938579,
        0.10796119],
       ...,
       [0.12110367, 0.07446129, 0.04558758, ..., 1.        ,
0.07085128,
        0.0196177 ],
       [0.07347567, 0.16773875, 0.04938579, ..., 0.07085128, 1.
,
        0.10602962],
       [0.04316046, 0.14263397, 0.10796119, ..., 0.0196177 ,
0.10602962,
        1.        ]])

similarity_scores.shape

(706, 706)

def recommend(book_name):
    # index fetch
    index = np.where(pt.index==book_name)[0][0]
    similar_items =
sorted(list(enumerate(similarity_scores[index])),key=lambda
x:x[1],reverse=True)[1:6]

    data = []
    for i in similar_items:
```

```python
        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
```

```python
recommend('1984')
```

```
[['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']]
```

```python
recommend("The Handmaid's Tale")
```

```
[['Bastard Out of Carolina',
  'Dorothy Allison',
  'http://images.amazon.com/images/P/0452269571.01.MZZZZZZZ.jpg'],
 ['A Civil Action',
  'JONATHAN HARR',
  'http://images.amazon.com/images/P/0679772677.01.MZZZZZZZ.jpg'],
 ['1984',
  'George Orwell',
  'http://images.amazon.com/images/P/0451524934.01.MZZZZZZZ.jpg'],
 ["The Kitchen God's Wife",
  'Amy Tan',
  'http://images.amazon.com/images/P/0399135782.01.MZZZZZZZ.jpg'],
 ['The Red Tent (Bestselling Backlist)',
  'Anita Diamant',
  'http://images.amazon.com/images/P/0312195516.01.MZZZZZZZ.jpg']]
```