

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

```
In [2]: books=pd.read_csv('Books.csv',sep=';',error_bad_lines=False,encoding='latin-1')
books.head()
```

```
b'Skipping line 6452: expected 8 fields, saw 9\nSkipping line 43667: expected 8
fields, saw 10\nSkipping line 51751: expected 8 fields, saw 9\n'
b'Skipping line 92038: expected 8 fields, saw 9\nSkipping line 104319: expected
8 fields, saw 9\nSkipping line 121768: expected 8 fields, saw 9\n'
b'Skipping line 144058: expected 8 fields, saw 9\nSkipping line 150789: expecte
d 8 fields, saw 9\nSkipping line 157128: expected 8 fields, saw 9\nSkipping lin
e 180189: expected 8 fields, saw 9\nSkipping line 185738: expected 8 fields, sa
w 9\n'
b'Skipping line 209388: expected 8 fields, saw 9\nSkipping line 220626: expecte
d 8 fields, saw 9\nSkipping line 227933: expected 8 fields, saw 11\nSkipping li
ne 228957: expected 8 fields, saw 10\nSkipping line 245933: expected 8 fields,
saw 9\nSkipping line 251296: expected 8 fields, saw 9\nSkipping line 259941: ex
pected 8 fields, saw 9\nSkipping line 261529: expected 8 fields, saw 9\n'
C:\Users\admin\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:316
5: DtypeWarning: Columns (3) have mixed types.Specify dtype option on import or
set low_memory=False.
```

```
has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[2]:

| | ISBN | Book-Title | Book- Author | Year-Of- Publication | Publisher | |
|---|------------|---|----------------------------|-------------------------|------------------------------|---|
| 0 | 0195153448 | Classical Mythology | Mark P. O. Morford | 2002 | Oxford University Press | http://images.amazon.com/images/P/0195153448.jpg |
| 1 | 0002005018 | Clara Callan | Richard Bruce Wright | 2001 | HarperFlamingo Canada | http://images.amazon.com/images/P/0002005018.jpg |
| 2 | 0060973129 | Decision in Normandy | Carlo D'Este | 1991 | HarperPerennial | http://images.amazon.com/images/P/0060973129.jpg |
| 3 | 0374157065 | Flu: The Story of the Great Influenza Pandemic... | Gina Bari Kolata | 1999 | Farrar Straus Giroux | http://images.amazon.com/images/P/0374157065.jpg |
| 4 | 0393045218 | The Mummies of Urumchi | E. J. W. Barber | 1999 | W. W. Norton & Company | http://images.amazon.com/images/P/0393045218.jpg |

```
In [3]: books.columns
```

```
Out[3]: Index(['ISBN', 'Book-Title', 'Book-Author', 'Year-Of-Publication', 'Publisher',
              'Image-URL-S', 'Image-URL-M', 'Image-URL-L'],
              dtype='object')
```

```
In [4]: books=books[['ISBN', 'Book-Title', 'Book-Author', 'Year-Of-Publication', 'Publisher']]
books.head(2)
```

Out[4]:

| | ISBN | Book-Title | Book-Author | Year-Of-Publication | Publisher |
|---|------------|---------------------|----------------------|---------------------|-------------------------|
| 0 | 0195153448 | Classical Mythology | Mark P. O. Morford | 2002 | Oxford University Press |
| 1 | 0002005018 | Clara Callan | Richard Bruce Wright | 2001 | HarperFlamingo Canada |

```
In [5]: books.rename(columns={'Book-Title':'title', 'Book-Author':'author', 'Year-Of-Publication':'year'})
books.head(2)
```

Out[5]:

| | ISBN | title | author | year | publisher |
|---|------------|---------------------|----------------------|------|-------------------------|
| 0 | 0195153448 | Classical Mythology | Mark P. O. Morford | 2002 | Oxford University Press |
| 1 | 0002005018 | Clara Callan | Richard Bruce Wright | 2001 | HarperFlamingo Canada |

```
In [6]: users=pd.read_csv('BX-users.csv', sep=';', error_bad_lines=False, encoding='latin-1')
users.head(2)
```

Out[6]:

| | User-ID | Location | Age |
|---|---------|---------------------------|------|
| 0 | 1 | nyc, new york, usa | NaN |
| 1 | 2 | stockton, california, usa | 18.0 |

```
In [7]: users.rename(columns={'User-ID':'user_id', 'Location':'location', 'Age':'age'}, inplace=True)
users.head(2)
```

Out[7]:

| | user_id | location | age |
|---|---------|---------------------------|------|
| 0 | 1 | nyc, new york, usa | NaN |
| 1 | 2 | stockton, california, usa | 18.0 |

```
In [8]: ratings=pd.read_csv('BX-Book-Ratings.csv', sep=';', error_bad_lines=False, encoding='latin-1')
ratings.head(2)
```

Out[8]:

| | User-ID | ISBN | Book-Rating |
|---|---------|------------|-------------|
| 0 | 276725 | 034545104X | 0 |
| 1 | 276726 | 0155061224 | 5 |

```
In [9]: ratings.rename(columns={'User-ID': 'user_id', 'ISBN': 'ISBN', 'Book-Rating': 'rating'})
ratings.head(2)
```

```
Out[9]:
```

| | user_id | ISBN | rating |
|---|---------|------------|--------|
| 0 | 276725 | 034545104X | 0 |
| 1 | 276726 | 0155061224 | 5 |

```
In [10]: print(books.shape)

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

```
(271360, 5)
(278858, 3)
(1149780, 3)
```

```
In [11]: ratings['user_id'].value_counts().shape
```

```
Out[11]: (105283,)
```

```
In [12]: x=ratings['user_id'].value_counts()>200
x[x].shape
```

```
Out[12]: (899,)
```

```
In [13]: y=x[x].index
y
```

```
Out[13]: Int64Index([ 11676, 198711, 153662,  98391,  35859, 212898, 278418,  76352,
                    110973, 235105,
                    ...,
                    116122,  28634, 188951,  59727, 155916, 274808,  73681,  9856,
                    268622,  44296],
                    dtype='int64', length=899)
```

```
In [14]: ratings=ratings[ratings['user_id'].isin(y)]

ratings.shape
```

```
Out[14]: (526356, 3)
```

In [15]: `ratings.head()`

Out[15]:

| | user_id | ISBN | rating |
|------|---------|------------|--------|
| 1456 | 277427 | 002542730X | 10 |
| 1457 | 277427 | 0026217457 | 0 |
| 1458 | 277427 | 003008685X | 8 |
| 1459 | 277427 | 0030615321 | 0 |
| 1460 | 277427 | 0060002050 | 0 |

In [16]: `ratings_with_books=ratings.merge(books,on='ISBN')`
`ratings_with_books.head(2)`

Out[16]:

| | user_id | ISBN | rating | title | author | year | publisher |
|---|---------|------------|--------|---|-------------------|------|-----------------------|
| 0 | 277427 | 002542730X | 10 | Politically Correct Bedtime Stories: Modern Ta... | James Finn Garner | 1994 | John Wiley & Sons Inc |
| 1 | 3363 | 002542730X | 0 | Politically Correct Bedtime Stories: Modern Ta... | James Finn Garner | 1994 | John Wiley & Sons Inc |

In [17]: `ratings_with_books.shape`

Out[17]: (487671, 7)

In [18]: `number_rating = ratings_with_books.groupby('title')['rating'].count().reset_index()`
`number_rating.rename(columns={'rating':'number of ratings'},inplace=True)`
`number_rating.head(3)`

Out[18]:

| | title | number of ratings |
|---|--|-------------------|
| 0 | A Light in the Storm: The Civil War Diary of ... | 2 |
| 1 | Always Have Popsicles | 1 |
| 2 | Apple Magic (The Collector's series) | 1 |

In [19]: `final_rating=ratings_with_books.merge(number_rating,on='title')`
`final_rating.shape`

Out[19]: (487671, 8)

```
In [20]: final_rating=final_rating[final_rating['number of ratings']>=50]
final_rating.head(2)
```

Out[20]:

| | user_id | ISBN | rating | title | author | year | publisher | number of ratings |
|---|---------|------------|--------|--|-------------------------|------|-----------------------------|-------------------------|
| 0 | 277427 | 002542730X | 10 | Politically Correct Bedtime Stories: Modern Ta... | James Finn Garner | 1994 | John Wiley & Sons Inc | 82 |
| 1 | 3363 | 002542730X | 0 | Politically Correct Bedtime Stories: Modern Ta... | James Finn Garner | 1994 | John Wiley & Sons Inc | 82 |

```
In [21]: final_rating.shape
```

Out[21]: (61853, 8)

```
In [22]: final_rating.drop_duplicates(['user_id', 'title'], inplace=True)
```

```
In [23]: final_rating.shape
```

Out[23]: (59850, 8)

```
In [24]: book_pivot=final_rating.pivot_table(columns='user_id',index='title',values='rating')
book_pivot
```

Out[24]:

| | user_id | 254 | 2276 | 2766 | 2977 | 3363 | 3757 | 4017 | 4385 | 6242 | 6251 | ... | 274004 | 274061 | 2 |
|--|---|-----|------|------|------|------|------|------|------|------|------|-----|--------|--------|---|
| | title | | | | | | | | | | | | | | |
| | 1984 | 9.0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | |
| | 1st to Die: A Novel | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | |
| | 2nd Chance | NaN | 10.0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | |
| | 4 Blondes | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.0 | ... | NaN | NaN | |
| | 84 Charing Cross Road | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | |
| | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| | Year of Wonders | NaN | NaN | NaN | 7.0 | NaN | NaN | NaN | NaN | 7.0 | NaN | ... | NaN | NaN | |
| | You Belong To Me | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | |
| | Zen and the Art of Motorcycle Maintenance: An Inquiry into Values | NaN | NaN | NaN | NaN | 0.0 | NaN | NaN | NaN | NaN | 0.0 | ... | NaN | NaN | |
| | Zoya | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | |
| | "Ol" Is for Outlaw" | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | |

742 rows × 888 columns



```
In [25]: book_pivot.fillna(0,inplace=True)
book_pivot
```

Out[25]:

| | user_id | 254 | 2276 | 2766 | 2977 | 3363 | 3757 | 4017 | 4385 | 6242 | 6251 | ... | 274004 | 274061 | 27 |
|--|---|-----|------|------|------|------|------|------|------|------|------|-----|--------|--------|----|
| | title | | | | | | | | | | | | | | |
| | 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 | |
| | 1st to Die: A Novel | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 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 | 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 | 0.0 | |
| | 84 Charing Cross Road | 0.0 | 0.0 | 0.0 | 0.0 | 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 | 7.0 | 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 | 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 | 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 | 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 | 0.0 | |

742 rows × 888 columns



```
In [26]: from scipy.sparse import csr_matrix
book_sparse=csr_matrix(book_pivot)
```

```
In [27]: type(book_sparse)
```

Out[27]: scipy.sparse.csr.csr_matrix

```
In [28]: from sklearn.neighbors import NearestNeighbors
model=NearestNeighbors(algorithm='brute')
```

```
In [29]: model.fit(book_sparse)
```

Out[29]: NearestNeighbors(algorithm='brute')

```
In [30]: distances,suggestions=model.kneighbors(book_pivot.iloc[237,:].values.reshape(1,-1
```

```
In [31]: distances
```

```
Out[31]: array([[ 0.          , 68.78953409, 69.5413546 , 72.64296249, 76.83098333,
        77.28518616]])
```

```
In [32]: suggestions
```

```
Out[32]: array([[237, 240, 238, 241, 184, 536]], dtype=int64)
```

```
In [33]: for i in range(len(suggestions)):
        print(book_pivot.index[suggestions[i]])
```

```
Index(['Harry Potter and the Chamber of Secrets (Book 2)',
      'Harry Potter and the Prisoner of Azkaban (Book 3)',
      'Harry Potter and the Goblet of Fire (Book 4)',
      'Harry Potter and the Sorcerer's Stone (Book 1)', 'Exclusive',
      'The Cradle Will Fall'],
      dtype='object', name='title')
```

```
In [34]: book_pivot.index[54]
```

```
Out[34]: 'Animal Farm'
```

```
In [35]: np.where(book_pivot.index=='Animal Farm')[0][0]
```

```
Out[35]: 54
```

```
In [36]: def recommend(name):
        book_id=np.where(book_pivot.index==name)[0][0]
        distances,suggestions=model.kneighbors(book_pivot.iloc[book_id,:].values.reshape(1,-1))

        for i in range(len(suggestions)):
            if i==0:
                print("THE SUGGESTION FOR BOOKS ",name,"ARE :")
            if not i:
                print(book_pivot.index[suggestions[i]])
```

```
In [37]: recommend('Exclusive')
```

```
THE SUGGESTION FOR BOOKS Exclusive ARE :
Index(['Exclusive', 'The Cradle Will Fall', 'The Long Road Home',
      'Jacob Have I Loved', 'No Safe Place', 'Eyes of a Child'],
      dtype='object', name='title')
```

```
In [ ]:
```


