

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

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
In [34]: books = pd.read_csv('booksdata/BX-Books.csv', sep=";", on_bad_lines='skip', encoding='latin-1')
```

```
C:\Users\manon\AppData\Local\Temp\ipykernel_21744\1408659485.py:1: DtypeWarning: Columns (3) have mixed types. Specify dtype option on import or set low_memory=False.  
books = pd.read_csv('booksdata/BX-Books.csv', sep=";", on_bad_lines='skip', encoding='latin-1')
```

```
In [36]: books.head(2)
```

```
Out[36]:
```

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-L
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/ir
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/ir



```
In [38]: books.iloc[109]['Image-URL-L']
```

```
Out[38]: 'http://images.amazon.com/images/P/0671864769.01.LZZZZZZZ.jpg'
```

```
In [40]: books.shape
```

```
Out[40]: (271360, 8)
```

```
In [42]: books.columns
```

```
Out[42]: Index(['ISBN', 'Book-Title', 'Book-Author', 'Year-Of-Publication', 'Publisher',  
               'Image-URL-S', 'Image-URL-M', 'Image-URL-L'],  
               dtype='object')
```

```
In [44]: books = books[['ISBN', 'Book-Title', 'Book-Author', 'Year-Of-Publication', 'Publishe
```

```
In [46]: books.head(2)
```

```
Out[46]:
```

	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/ir
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/ir



```
In [50]: users = pd.read_csv('booksdata/BX-Users.csv', sep=";", on_bad_lines='skip', encoding='utf-8')
```

```
In [52]: users.head(2)
```

```
Out[52]:
```

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

```
In [54]: users.shape
```

```
Out[54]: (278858, 3)
```

```
In [58]: ratings = pd.read_csv('booksdata/BX-Book-Ratings.csv', sep=";", on_bad_lines='skip')
```

```
In [60]: ratings.head(2)
```

```
Out[60]:
```

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

```
In [62]: ratings.shape
```

```
Out[62]: (1149780, 3)
```

```
In [64]: ratings.rename(columns={"User-ID":'user_id', 'Book-Rating':'rating'},inplace=True)
```

```
In [66]: ratings.head(2)
```

```
Out[66]:
```

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

```
In [68]: print(books.shape, users.shape, ratings.shape, sep='\n')
```

```
(271360, 6)
(278858, 3)
(1149780, 3)
```

```
In [70]: ratings['user_id'].value_counts()
```

```
Out[70]: user_id
11676      13602
198711     7550
153662     6109
98391      5891
35859      5850
...
116180      1
116166      1
116154      1
116137      1
276723      1
Name: count, Length: 105283, dtype: int64
```

```
In [72]: ratings['user_id'].value_counts().shape
```

```
Out[72]: (105283,)
```

```
In [74]: ratings['user_id'].unique().shape
```

```
Out[74]: (105283,)
```

```
In [76]: x = ratings['user_id'].value_counts() > 200
```

```
In [78]: x[x].shape
```

```
Out[78]: (899,)
```

```
In [80]: y=x[x].index
```

```
In [82]: y
```

```
Out[82]: Index([ 11676, 198711, 153662, 98391, 35859, 212898, 278418, 76352, 110973,
235105,
...
260183, 73681, 44296, 155916, 9856, 274808, 28634, 59727, 268622,
188951],
dtype='int64', name='user_id', length=899)
```

```
In [84]: ratings = ratings[ratings['user_id'].isin(y)]
```

```
In [86]: ratings.head(3)
```

```
Out[86]:
```

	user_id	ISBN	rating
1456	277427	002542730X	10
1457	277427	0026217457	0
1458	277427	003008685X	8

```
In [88]: ratings.shape
```

```
Out[88]: (526356, 3)
```

```
In [90]: ratings_with_books = ratings.merge(books, on='ISBN')
```

```
In [110...]: ratings_with_books.rename(columns={"User-ID":'user_id',
                                             'Book-Rating':'rating',
                                             'Book-Title':'title',
                                             'Book-Author':'author',
                                             'Year-Of-Publication':'year',
                                             'Publisher':'publisher',
                                             'Image-URL-L':'image_url'},inplace=True)
```

```
In [112...]: ratings_with_books.head()
```

```
Out[112...]:
```

	user_id	ISBN	rating	title	author	publisher	publisher	
0	277427	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc	http://images.ai
1	277427	0026217457	0	Vegetarian Times Complete Cookbook	Lucy Moll	1995	John Wiley & Sons	http://images.ai
2	277427	003008685X	8	Pioneers	James Fenimore Cooper	1974	Thomson Learning	http://images.ai
3	277427	0030615321	0	Ask for May, Settle for June (A Doonesbury book)	G. B. Trudeau	1982	Henry Holt & Co	http://images.ai
4	277427	0060002050	0	On a Wicked Dawn (Cynster Novels)	Stephanie Laurens	2002	Avon Books	http://images.ai

```
In [114... ratings_with_books.shape
```

```
Out[114... (487671, 8)
```

```
In [116... number_rating = ratings_with_books.groupby('title')['rating'].count().reset_index()
```

```
In [118... number_rating.head()
```

```
Out[118...

|   | title                                            | rating |
|---|--------------------------------------------------|--------|
| 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      |
| 3 | Beyond IBM: Leadership Marketing and Finance ... | 1      |
| 4 | Clifford Visita El Hospital (Clifford El Gran... | 1      |


```

```
In [120... number_rating.rename(columns={'rating':'num_of_rating'}, inplace=True)
```

```
In [122... number_rating.head()
```

```
Out[122...

|   | title                                            | num_of_rating |
|---|--------------------------------------------------|---------------|
| 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             |
| 3 | Beyond IBM: Leadership Marketing and Finance ... | 1             |
| 4 | Clifford Visita El Hospital (Clifford El Gran... | 1             |


```

```
In [124... final_rating = ratings_with_books.merge(number_rating, on='title')
```

```
In [126... final_rating.head()
```

Out[126...]

	user_id	ISBN	rating	title	author	publisher	publisher
0	277427	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc http://images.ai
1	277427	0026217457	0	Vegetarian Times Complete Cookbook	Lucy Moll	1995	John Wiley & Sons http://images.a
2	277427	003008685X	8	Pioneers	James Fenimore Cooper	1974	Thomson Learning http://images.ai
3	277427	0030615321	0	Ask for May, Settle for June (A Doonesbury book)	G. B. Trudeau	1982	Henry Holt & Co http://images.a
4	277427	0060002050	0	On a Wicked Dawn (Cynster Novels)	Stephanie Laurens	2002	Avon Books http://images.a



In [128...]

final_rating.shape

Out[128...]

(487671, 9)

In [130...]

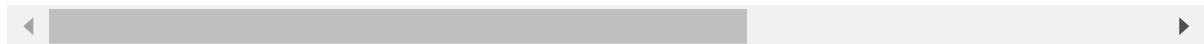
final_rating = final_rating[final_rating['num_of_rating'] >= 50]

In [132...]

final_rating.head()

```
Out[132...]
```

	<code>user_id</code>	<code>ISBN</code>	<code>rating</code>	<code>title</code>	<code>author</code>	<code>publisher</code>	<code>publisher</code>
0	277427	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	James Finn Garner	1994	John Wiley & Sons Inc
13	277427	0060930535	0	The Poisonwood Bible: A Novel	Barbara Kingsolver	1999	Perennial
15	277427	0060934417	0	Bel Canto: A Novel	Ann Patchett	2002	Perennial
18	277427	0061009059	9	One for the Money (Stephanie Plum Novels (Pape...	Janet Evanovich	1995	HarperTorch
24	277427	006440188X	0	The Secret Garden	Frances Hodgson Burnett	1998	HarperTrophy



```
In [134...]
```

```
final_rating.shape
```

```
Out[134...]
```

```
(61853, 9)
```

```
In [140...]
```

```
final_rating.drop_duplicates(['user_id', 'title'])
```

Out[140...]

	user_id	ISBN	rating	title	author	publisher	publisher
0	277427	002542730X	10	Politically Correct Bedtime Stories: Modern Tales	James Finn Garner	1994	John Wiley & Sons Inc
13	277427	0060930535	0	The Poisonwood Bible: A Novel	Barbara Kingsolver	1999	Perennial
15	277427	0060934417	0	Bel Canto: A Novel	Ann Patchett	2002	Perennial
18	277427	0061009059	9	One for the Money (Stephanie Plum Novels (Paperback))	Janet Evanovich	1995	HarperTorch
24	277427	006440188X	0	The Secret Garden	Frances Hodgson Burnett	1998	HarperTrophy
...
487505	275970	1400031354	0	Tears of the Giraffe (No.1 Ladies Detective Agency)	Alexander McCall Smith	2002	Anchor
487506	275970	1400031362	0	Morality for Beautiful Girls (No.1 Ladies Detective Agency)	Alexander McCall Smith	2002	Anchor
487579	275970	1573229725	0	Fingersmith	Sarah Waters	2002	Riverhead Books
487618	275970	1586210661	9	Me Talk Pretty One Day	David Sedaris	2001	Time Warner Audio Major
487619	275970	1586212230	0	Naked	David Sedaris	2001	Time Warner Audio Major

59850 rows × 9 columns



In [148...]

final_rating.shape

```
Out[148... (59850, 9)
```

```
In [150... book_pivot = final_rating.pivot_table(columns='user_id', index='title', values= 'ra
```

```
In [152... book_pivot
```

```
Out[152... user_id 254 2276 2766 2977 3363 3757 4017 4385 6242 6251 ... 274004
```

title													
1984	9.0	NaN	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	NaN
2nd Chance	NaN	10.0	NaN										
4 Blondes	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	...	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	NaN	7.0	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
\O\ Is for Outlaw"	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN

742 rows × 888 columns



```
In [154... book_pivot.shape
```

```
Out[154... (742, 888)
```

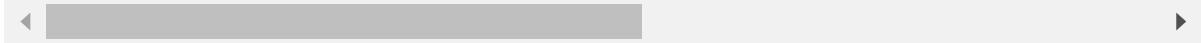
```
In [156... book_pivot.fillna(0, inplace=True)
```

```
In [158... book_pivot
```

Out[158...]

	user_id	254	2276	2766	2977	3363	3757	4017	4385	6242	6251	...	274004	
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	...	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	...	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	...	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	...	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	...	0.0
...
Year of Wonders	0.0	0.0	0.0	7.0	0.0	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	...	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	...	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	...	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	...	0.0

742 rows × 888 columns



In [160...]

```
from scipy.sparse import csr_matrix
```

In [162...]

```
book_sparse = csr_matrix(book_pivot)
```

In [164...]

```
type(book_sparse)
```

Out[164...]

```
scipy.sparse._csr.csr_matrix
```

In [166...]

```
from sklearn.neighbors import NearestNeighbors
model = NearestNeighbors(algorithm= 'brute')
```

In [168...]

```
model.fit(book_sparse)
```

```
Out[168...]
```

▼ NearestNeighbors 

```
NearestNeighbors(algorithm='brute')
```

```
In [170...]
```

```
distance, suggestion = model.kneighbors(book_pivot.iloc[237,:].values.reshape(1,-1))
```

```
In [172...]
```

```
distance
```

```
Out[172...]
```

```
array([[ 0.          , 67.75691847, 68.05145112, 72.277244  , 75.81556568,
        76.30203143]])
```

```
In [174...]
```

```
suggestion
```

```
Out[174...]
```

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

```
In [176...]
```

```
book_pivot.iloc[241,:]
```

```
Out[176...]
```

```
user_id
```

```
254      9.0
2276     0.0
2766     0.0
2977     0.0
3363     0.0
...
275970    9.0
277427    0.0
277478    0.0
277639    0.0
278418    0.0
```

```
Name: Harry Potter and the Sorcerer's Stone (Book 1), Length: 888, dtype: float64
```

```
In [178...]
```

```
for i in range(len(suggestion)):
    print(book_pivot.index[suggestion[i]])
```

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

```
In [180...]
```

```
book_pivot.index[3]
```

```
Out[180...]
```

```
'4 Blondes'
```

```
In [182...]
```

```
book_names = book_pivot.index
```

```
In [184...]
```

```
book_names[2]
```

```
Out[184...]
```

```
'2nd Chance'
```

```
In [186...]
```

```
np.where(book_pivot.index == '4 Blondes')[0][0]
```

```
Out[186... 3

In [188... ids = np.where(final_rating['title'] == "Harry Potter and the Chamber of Secrets (Book 2)")[0]
In [190... final_rating.iloc[ids]['image_url']
Out[190... 'http://images.amazon.com/images/P/0439064872.01.LZZZZZZZ.jpg'

In [192... book_name = []
for book_id in suggestion:
    book_name.append(book_pivot.index[book_id])
In [194... book_name[0]
Out[194... Index(['Harry Potter and the Chamber of Secrets (Book 2)', 'Harry Potter and the Goblet of Fire (Book 4)', 'Harry Potter and the Prisoner of Azkaban (Book 3)', 'Harry Potter and the Sorcerer's Stone (Book 1)', 'Exclusive', 'The Cradle Will Fall'], dtype='object', name='title')

In [196... ids_index = []
for name in book_name[0]:
    ids = np.where(final_rating['title'] == name)[0][0]
    ids_index.append(ids)
In [198... for idx in ids_index:
    url = final_rating.iloc[idx]['image_url']
    print(url)
http://images.amazon.com/images/P/0439064872.01.LZZZZZZZ.jpg
http://images.amazon.com/images/P/0439139597.01.LZZZZZZZ.jpg
http://images.amazon.com/images/P/0439136369.01.LZZZZZZZ.jpg
http://images.amazon.com/images/P/043936213X.01.LZZZZZZZ.jpg
http://images.amazon.com/images/P/0446604232.01.LZZZZZZZ.jpg
http://images.amazon.com/images/P/0440115450.01.LZZZZZZZ.jpg

In [202... import pickle
pickle.dump(model, open('artifacts/model.pkl', 'wb'))
pickle.dump(book_names, open('artifacts/book_names.pkl', 'wb'))
pickle.dump(final_rating, open('artifacts/final_rating.pkl', 'wb'))
pickle.dump(book_pivot, open('artifacts/book_pivot.pkl', 'wb'))

In [204... def recommend_book(book_name):
    book_id = np.where(book_pivot.index == book_name)[0][0]
    distance, suggestion = model.kneighbors(book_pivot.iloc[book_id, :].values.reshape(1, -1))

    for i in range(len(suggestion)):
        books = book_pivot.index[suggestion[i]]
        for j in books:
            if j == book_name:
                print(f"You searched '{book_name}'\n")
                print("The suggestion books are: \n")
```

```
    else:  
        print(j)
```

In [206]:

```
book_name = "Harry Potter and the Chamber of Secrets (Book 2)"  
recommend_book(book_name)
```

You searched 'Harry Potter and the Chamber of Secrets (Book 2)'

The suggestion books are:

```
Harry Potter and the Goblet of Fire (Book 4)  
Harry Potter and the Prisoner of Azkaban (Book 3)  
Harry Potter and the Sorcerer's Stone (Book 1)  
Exclusive  
The Cradle Will Fall
```

In []: