

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: import pickle
```

Loading Dataset

```
In [3]: book=pd.read_csv("Books.csv")
```

C:\Users\shaw3\AppData\Local\Temp\ipykernel_10152\3971472040.py:1: DtypeWarning: Columns (3) have mixed types. Specify dtype option on import or set low_memory=False.

```
book=pd.read_csv("Books.csv")
```

```
In [4]: book
```

Out[4]:

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	h
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	h
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	h
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	h
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	h
...	
271355	0440400988	There's a Bat in Bunk Five	Paula Danziger	1988	Random House Childrens Pub (Mm)	h
271356	0525447644	From One to One Hundred	Teri Sloat	1991	Dutton Books	h
271357	006008667X	Lily Dale : The True Story of the Town that Ta...	Christine Wicker	2004	HarperSanFrancisco	h
271358	0192126040	Republic (World's Classics)	Plato	1996	Oxford University Press	h
271359	0767409752	A Guided Tour of Rene Descartes' Meditations O...	Christopher Biffle	2000	McGraw-Hill Humanities/Social Sciences/Languages	h

271360 rows × 8 columns

In [5]: `book.head()`

Out[5]:

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

In [6]: `book.describe()`

Out[6]:

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	
count	271360	271360	271359	271360	271358	
unique	271360	242135	102023	202	16807	
top	0195153448	Selected Poems	Agatha Christie	2002	Harlequin	http://images.ama:
freq	1	27	632	13903	7535	

In [7]: `book.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271360 entries, 0 to 271359
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   ISBN                  271360 non-null object
1   Book-Title            271360 non-null object
2   Book-Author           271359 non-null object
3   Year-Of-Publication    271360 non-null object
4   Publisher              271358 non-null object
5   Image-URL-S            271360 non-null object
6   Image-URL-M            271360 non-null object
7   Image-URL-L            271357 non-null object
dtypes: object(8)
memory usage: 16.6+ MB
```

In [8]: `book.count()`

```
Out[8]: ISBN                271360
        Book-Title          271360
        Book-Author         271359
        Year-Of-Publication  271360
        Publisher            271358
        Image-URL-S          271360
        Image-URL-M          271360
        Image-URL-L          271357
        dtype: int64
```

```
In [9]: book.columns
```

```
Out[9]: Index(['ISBN', 'Book-Title', 'Book-Author', 'Year-Of-Publication', 'Publish
er',
              'Image-URL-S', 'Image-URL-M', 'Image-URL-L'],
              dtype='object')
```

```
In [10]: book.shape
```

```
Out[10]: (271360, 8)
```

```
In [11]: print(f"Column types in book:\n{book.dtypes}")
```

```
Column types in book:
ISBN                object
Book-Title          object
Book-Author         object
Year-Of-Publication object
Publisher            object
Image-URL-S          object
Image-URL-M          object
Image-URL-L          object
dtype: object
```

```
In [12]: user = pd.read_csv('Users ok.csv')
```

```
In [13]: user
```

Out[13]:

	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 × 3 columns

In [14]: `user.head()`

Out[14]:

	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 [15]: `user.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 278858 entries, 0 to 278857
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   User-ID     278858 non-null  int64
1   Location    278858 non-null  object
2   Age         168096 non-null  float64
dtypes: float64(1), int64(1), object(1)
memory usage: 6.4+ MB
```

In [16]: `user.describe()`

```
Out[16]:
```

	User-ID	Age
count	278858.00000	168096.000000
mean	139429.50000	34.751434
std	80499.51502	14.428097
min	1.00000	0.000000
25%	69715.25000	24.000000
50%	139429.50000	32.000000
75%	209143.75000	44.000000
max	278858.00000	244.000000

```
In [17]: user.shape
```

```
Out[17]: (278858, 3)
```

```
In [18]: user.count()
```

```
Out[18]: User-ID      278858
Location    278858
Age         168096
dtype: int64
```

```
In [19]: user.columns
```

```
Out[19]: Index(['User-ID', 'Location', 'Age'], dtype='object')
```

CHECK IF NULL VALUE PRESENT OR NOT

```
In [20]: user.isnull().sum()
```

```
Out[20]: User-ID      0
Location    0
Age        110762
dtype: int64
```

```
In [21]: rating=pd.read_csv("Ratings.csv")
```

```
In [22]: rating
```

```
Out[22]:
```

	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 × 3 columns

```
In [23]: rating.head()
```

```
Out[23]:
```

	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

```
In [24]: rating.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1149780 entries, 0 to 1149779
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   User-ID     1149780 non-null  int64
1   ISBN        1149780 non-null  object
2   Book-Rating 1149780 non-null  int64
dtypes: int64(2), object(1)
memory usage: 26.3+ MB
```

```
In [25]: rating.describe()
```

```
Out[25]:
```

	User-ID	Book-Rating
count	1.149780e+06	1.149780e+06
mean	1.403864e+05	2.866950e+00
std	8.056228e+04	3.854184e+00
min	2.000000e+00	0.000000e+00
25%	7.034500e+04	0.000000e+00
50%	1.410100e+05	0.000000e+00
75%	2.110280e+05	7.000000e+00
max	2.788540e+05	1.000000e+01

```
In [26]: rating.count()
```

```
Out[26]: User-ID      1149780
ISBN          1149780
Book-Rating   1149780
dtype: int64
```

```
In [27]: rating.columns
```

```
Out[27]: Index(['User-ID', 'ISBN', 'Book-Rating'], dtype='object')
```

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

```
Out[28]: (1149780, 3)
```

```
In [29]: rating.isnull().sum()
```

```
Out[29]: User-ID      0
ISBN          0
Book-Rating   0
dtype: int64
```

```
In [30]: user['Location'].value_counts()
```

```
Out[30]: london, england, united kingdom      2506
toronto, ontario, canada                     2250
sydney, new south wales, australia           1744
melbourne, victoria, australia               1708
portland, oregon, usa                        1629
...
gneisenaustr. 3, nordrhein-westfalen, germany      1
-, bayern, germany                                1
bouchegouf, guelma, algeria                      1
indiana and philadelphia, pennsylvania, usa       1
tacoma, washington, united kingdom                1
Name: Location, Length: 57339, dtype: int64
```

```
In [31]: user[['City', 'State', 'Country']] = user['Location'].str.split(',', expand
```


check duplicate columns

```
In [32]: new_books = book.drop_duplicates('Book-Title')
```

```
In [33]: new_books
```

Out[33]:

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

242135 rows × 8 columns

```
In [34]: new_books.shape
```

```
Out[34]: (242135, 8)
```

```
In [35]: ratings_with_name = rating.merge(new_books, on='ISBN')
```

```
In [36]: ratings_with_name.head()
```

```
Out[36]:
```

	User-ID	ISBN	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher	
0	276725	034545104X	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://im
1	2313	034545104X	5	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://im
2	6543	034545104X	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://im
3	8680	034545104X	5	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://im
4	10314	034545104X	9	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	http://im

```
In [37]: ratings_with_name.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 883079 entries, 0 to 883078
Data columns (total 10 columns):
#   Column                      Non-Null Count  Dtype
---  -
0   User-ID                     883079 non-null  int64
1   ISBN                        883079 non-null  object
2   Book-Rating                 883079 non-null  int64
3   Book-Title                  883079 non-null  object
4   Book-Author                 883078 non-null  object
5   Year-Of-Publication         883079 non-null  object
6   Publisher                   883078 non-null  object
7   Image-URL-S                 883079 non-null  object
8   Image-URL-M                 883079 non-null  object
9   Image-URL-L                 883075 non-null  object
dtypes: int64(2), object(8)
memory usage: 74.1+ MB
```

```
In [38]: ratings_with_name.describe()
```

Out[38]:

	User-ID	Book-Rating
count	883079.000000	883079.000000
mean	140001.288996	2.832366
std	80927.312913	3.850714
min	2.000000	0.000000
25%	69697.000000	0.000000
50%	140069.000000	0.000000
75%	211351.000000	7.000000
max	278854.000000	10.000000

In [39]: ratings_with_name.shape

Out[39]: (883079, 10)

In [40]: ratings_with_name.count

```

Out[40]: <bound method DataFrame.count of
\
0      276725  034545104X      0
1      2313   034545104X      5
2      6543   034545104X      0
3      8680   034545104X      5
4     10314   034545104X      9
...      ...      ...      ...
883074  276688  0425150526      0
883075  276688  0449907422      0
883076  276690  0590907301      0
883077  276704  0679752714      0
883078  276704  0806917695      5

                                     Book-Title      Book-Author
\
0      Flesh Tones: A Novel      M. J. Rose
1      Flesh Tones: A Novel      M. J. Rose
2      Flesh Tones: A Novel      M. J. Rose
3      Flesh Tones: A Novel      M. J. Rose
4      Flesh Tones: A Novel      M. J. Rose
...      ...      ...
883074      Death Crosses the Border      Janice Steinberg
883075      Jazz Funeral: A Skip Langdon Novel      Julie Smith
883076  Triplet Trouble and the Class Trip (Triplet Tr...      Debbie Dadey
883077  A Desert of Pure Feeling (Vintage Contemporaries)      Judith Freeman
883078  Perplexing Lateral Thinking Puzzles: Scholasti...      Paul Sloane

      Year-Of-Publication      Publisher \
0      2002      Ballantine Books
1      2002      Ballantine Books
2      2002      Ballantine Books
3      2002      Ballantine Books
4      2002      Ballantine Books
...      ...      ...
883074      1995      Berkley Publishing Group
883075      1993      Ballantine Books
883076      1997      Apple
883077      1997      Vintage Books USA
883078      1997      Sterling Publishing

                                     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...
...      ...
883074  http://images.amazon.com/images/P/0425150526.0...
883075  http://images.amazon.com/images/P/0449907422.0...
883076  http://images.amazon.com/images/P/0590907301.0...
883077  http://images.amazon.com/images/P/0679752714.0...
883078  http://images.amazon.com/images/P/0806917695.0...

                                     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...
...
883074 http://images.amazon.com/images/P/0425150526.0...
883075 http://images.amazon.com/images/P/0449907422.0...
883076 http://images.amazon.com/images/P/0590907301.0...
883077 http://images.amazon.com/images/P/0679752714.0...
883078 http://images.amazon.com/images/P/0806917695.0...

```

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...
...
883074 http://images.amazon.com/images/P/0425150526.0...
883075 http://images.amazon.com/images/P/0449907422.0...
883076 http://images.amazon.com/images/P/0590907301.0...
883077 http://images.amazon.com/images/P/0679752714.0...
883078 http://images.amazon.com/images/P/0806917695.0...

```

[883079 rows x 10 columns]>

```
In [41]: ratings_with_name.drop(['ISBN', 'Image-URL-S', 'Image-URL-M', 'Image-URL-L'])
```

```
In [42]: ratings_with_name.head()
```

```
Out[42]:
```

	User-ID	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher
0	276725	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
1	2313	5	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
2	6543	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
3	8680	5	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
4	10314	9	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books

```
In [43]: users_ratings_matrix = ratings_with_name.merge(user, on='User-ID')
```

```
In [44]: users_ratings_matrix.head()
```

	User-ID	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher	Location
0	276725	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	texas,
1	2313	5	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books	cincinnati, ohio,
2	2313	8	In Cold Blood (Vintage International)	TRUMAN CAPOTE	1994	Vintage	cincinnati, ohio,
3	2313	9	Divine Secrets of the Ya-Ya Sisterhood : A Novel	Rebecca Wells	1996	HarperCollins	cincinnati, ohio,
4	2313	5	The Mistress of Spices	Chitra Banerjee Divakaruni	1998	Anchor Books/Doubleday	cincinnati, ohio,

In [45]: `users_ratings_matrix.shape`

Out[45]: (883079, 11)

In [46]: `users_ratings_matrix.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 883079 entries, 0 to 883078
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   User-ID                               883079 non-null  int64
1   Book-Rating                           883079 non-null  int64
2   Book-Title                            883079 non-null  object
3   Book-Author                           883078 non-null  object
4   Year-Of-Publication                    883079 non-null  object
5   Publisher                              883078 non-null  object
6   Location                               883079 non-null  object
7   Age                                    646933 non-null  float64
8   City                                   883079 non-null  object
9   State                                  883079 non-null  object
10  Country                                883079 non-null  object
dtypes: float64(1), int64(2), object(8)
memory usage: 80.8+ MB
```

In [47]: `users_ratings_matrix.describe`

```

Out[47]: <bound method NDFrame.describe of          User-ID  Book-Rating  \
0          276725          0
1          2313          5
2          2313          8
3          2313          9
4          2313          5
...          ...          ...
883074    276618          5
883075    276647          0
883076    276647         10
883077    276660          8
883078    276672          5

```

```

                                Book-Title  \
0                                Flesh Tones: A Novel
1                                Flesh Tones: A Novel
2                                In Cold Blood (Vintage International)
3                                Divine Secrets of the Ya-Ya Sisterhood : A Novel
4                                The Mistress of Spices
...                                ...
883074    Ludwig Marum: Briefe aus dem Konzentrationslag...
883075    Christmas With Anne and Other Holiday Stories:...
883076    Heaven (Coretta Scott King Author Award Winner)
883077                                ROBOT RACE (MICRO ADV 6)
883078                                Dry Rot And Daffodils/Ne

```

```

                                Book-Author  Year-Of-Publication  \
0                                M. J. Rose          2002
1                                M. J. Rose          2002
2                                TRUMAN CAPOTE        1994
3                                Rebecca Wells        1996
4                                Chitra Banerjee Divakaruni  1998
...                                ...
883074                                Ludwig Marum          1984
883075                                L. M. Montgomery        2001
883076                                Angela Johnson          1998
883077                                David Antony Kroft        1985
883078                                Mary MacKie            2000

```

```

                                Publisher  \
0                                Ballantine Books
1                                Ballantine Books
2                                Vintage
3                                HarperCollins
4                                Anchor Books/Doubleday
...                                ...
883074                                C.F. MÃ¼ller
883075                                Starfire
883076    Simon & Schuster Children's Publishing
883077                                HarperCollins Publishers
883078                                Trafalgar Square

```

```

                                Location  Age          City          Stat
e  \
0                                tyler, texas, usa    NaN          tyler          texa
s

```

```

1          cincinnati, ohio, usa  23.0          cincinnati      ohi
0
2          cincinnati, ohio, usa  23.0          cincinnati      ohi
0
3          cincinnati, ohio, usa  23.0          cincinnati      ohi
0
4          cincinnati, ohio, usa  23.0          cincinnati      ohi
0
...
...
883074      stuttgart, \n/a\"., germany"  NaN          stuttgart      \n/a
\".
883075  arlington heights, illinois, usa  13.0  arlington heights  illinoi
s
883076  arlington heights, illinois, usa  13.0  arlington heights  illinoi
s
883077          singapore, n/a, singapore  15.0          singapore      n/
a
883078      newport, gwent, united kingdom  NaN          newport      gwen
t

```

```

          Country
0          usa
1          usa
2          usa
3          usa
4          usa
...
...
883074      germany"
883075          usa
883076          usa
883077      singapore
883078  united kingdom

```

[883079 rows x 11 columns]>

```
In [48]: users_ratings_matrix.isna().sum()
```

```

Out[48]: User-ID          0
Book-Rating          0
Book-Title          0
Book-Author          1
Year-Of-Publication  0
Publisher            1
Location             0
Age                 236146
City                 0
State                0
Country              0
dtype: int64

```

Top 50 Books based on Popularity


```
In [51]: ratings_with_books = rating.merge(book,on='ISBN')
ratings_with_books.shape
```

```
Out[51]: (1031136, 10)
```

```
In [52]: ratings_with_books
```

```
Out[52]:
```

	User-ID	ISBN	Book-Rating	Book-Title	Book-Author	Year-Of-Publication
0	276725	034545104X	0	Flesh Tones: A Novel	M. J. Rose	2002
1	2313	034545104X	5	Flesh Tones: A Novel	M. J. Rose	2002
2	6543	034545104X	0	Flesh Tones: A Novel	M. J. Rose	2002
3	8680	034545104X	5	Flesh Tones: A Novel	M. J. Rose	2002
4	10314	034545104X	9	Flesh Tones: A Novel	M. J. Rose	2002
...
1031131	276688	0517145553	0	Mostly Harmless	Douglas Adams	1995
1031132	276688	1575660792	7	Gray Matter	Shirley Kennett	1996
1031133	276690	0590907301	0	Triplet Trouble and the Class Trip (Triplet Tr...	Debbie Dadey	1997
1031134	276704	0679752714	0	A Desert of Pure Feeling (Vintage Contemporaries)	Judith Freeman	1997
1031135	276704	0806917695	5	Perplexing Lateral Thinking Puzzles: Scholasti...	Paul Sloane	1997

1031136 rows × 10 columns

```
In [108... ratings_with_books.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1031136 entries, 0 to 1031135
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   User-ID                               1031136 non-null  int64
1   ISBN                                  1031136 non-null  object
2   Book-Rating                           1031136 non-null  int64
3   Book-Title                            1031136 non-null  object
4   Book-Author                           1031135 non-null  object
5   Year-Of-Publication                   1031136 non-null  object
6   Publisher                             1031134 non-null  object
7   Image-URL-S                           1031136 non-null  object
8   Image-URL-M                           1031136 non-null  object
9   Image-URL-L                           1031132 non-null  object
dtypes: int64(2), object(8)
memory usage: 86.5+ MB

```

```
In [109... ratings_with_books.columns
```

```
Out[109... Index(['User-ID', 'ISBN', 'Book-Rating', 'Book-Title', 'Book-Author',
        'Year-Of-Publication', 'Publisher', 'Image-URL-S', 'Image-URL-M',
        'Image-URL-L'],
        dtype='object')
```

```
In [54]: num_rating = ratings_with_name.groupby('Book-Title').count()['Book-Rating'].
num_rating.rename(columns={'Book-Rating': 'num_ratings'}, inplace=True)
num_rating
```

```
Out[54]:
```

	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
...
241056	Ã?Â?piraten.	2
241057	Ã?Â?rger mit Produkt X. Roman.	4
241058	Ã?Â?sterlich leben.	1
241059	Ã?Â?stlich der Berge.	3
241060	Ã?Â?thique en toc	2

241061 rows × 2 columns

```
In [55]: avg_rating = ratings_with_name.groupby('Book-Title')['Book-Rating'].mean().r
avg_rating.rename(columns={'Book-Rating': 'avg_rating'}, inplace=True)
avg_rating
```

Out[55]:

	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
...
241056	Ã?Ã?piraten.	0.000000
241057	Ã?Ã?rger mit Produkt X. Roman.	5.250000
241058	Ã?Ã?sterlich leben.	7.000000
241059	Ã?Ã?stlich der Berge.	2.666667
241060	Ã?Ã?thique en toc	4.000000

241061 rows × 2 columns

```
In [57]: popular = num_rating.merge(avg_rating,on='Book-Title')
popular
```

Out[57]:

	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
...
241056	Ã?Ã?piraten.	2	0.000000
241057	Ã?Ã?rger mit Produkt X. Roman.	4	5.250000
241058	Ã?Ã?sterlich leben.	1	7.000000
241059	Ã?Ã?stlich der Berge.	3	2.666667
241060	Ã?Ã?thique en toc	2	4.000000

241061 rows × 3 columns

```
In [58]: popular = popular[popular['num_ratings']>=250].sort_values('avg_rating',asce
```

```
In [59]: popular
```

Out[59]:

	Book-Title	num_ratings	avg_rating
80421	Harry Potter and the Order of the Phoenix (Boo...	334	5.571856
191602	The Hobbit : The Enchanting Prelude to The Lor...	281	5.007117
219731	To Kill a Mockingbird	389	4.920308
80440	Harry Potter and the Sorcerer's Stone (Harry P...	571	4.900175
80409	Harry Potter and the Chamber of Secrets (Book 2)	351	4.729345
183564	The Da Vinci Code	883	4.652322
180547	The Catcher in the Rye	403	4.635236
187871	The Five People You Meet in Heaven	427	4.543326
187368	The Fellowship of the Ring (The Lord of the Ri...	257	4.505837
196316	The Lovely Bones: A Novel	1295	4.468726
144158	Prodigal Summer: A Novel	253	4.450593
223125	Tuesdays with Morrie: An Old Man, a Young Man,...	391	4.406650
168711	Stupid White Men ...and Other Sorry Excuses fo...	283	4.356890
204377	The Red Tent (Bestselling Backlist)	723	4.334716
206492	The Secret Life of Bees	615	4.219512
74747	Girl with a Pearl Earring	526	4.218631
129373	Nickel and Dime: On (Not) Getting By in America	267	4.142322
8434	ANGELA'S ASHES	269	4.115242
233360	Where the Heart Is (Oprah's Book Club (Paperba...	585	4.105983
2281	A Child Called \It\": One Child's Courage to S...	262	4.091603
76340	Good in Bed	470	4.040426
64928	Fast Food Nation: The Dark Side of the All- Ame...	321	4.037383
105772	Left Behind: A Novel of the Earth's Last Days ...	297	4.026936
12699	American Gods	302	4.006623
5664	A Prayer for Owen Meany	412	3.781553
212060	The Vampire Lestat (Vampire Chronicles, Book II)	301	3.777409

	Book-Title	num_ratings	avg_rating
233841	White Oleander : A Novel (Oprah's Book Club)	356	3.772472
234730	Wicked: The Life and Times of the Wicked Witch...	326	3.766871
219339	Timeline	407	3.761671
14391	Angels & Demons	586	3.718430
21753	Bel Canto: A Novel	350	3.702857
118232	Me Talk Pretty One Day	340	3.691176
19779	Balzac and the Little Chinese Seamstress : A N...	283	3.657244
134487	One for the Money (Stephanie Plum Novels (Pape...	291	3.656357
38811	Confessions of a Shopaholic (Summer Display Op...	350	3.634286
112927	Lucky : A Memoir	301	3.621262
148031	Red Dragon	365	3.619178
202540	The Poisonwood Bible: A Novel	494	3.609312
199227	The Nanny Diaries: A Novel	474	3.597046
202234	The Pilot's Wife : A Novel	568	3.593310
159293	She's Come Undone (Oprah's Book Club)	468	3.570513
93205	Interview with the Vampire	506	3.535573
818	1st to Die: A Novel	391	3.498721
159292	She's Come Undone (Oprah's Book Club (Paperback))	446	3.495516
223850	Two for the Dough	267	3.464419
60163	Empire Falls	321	3.454829
190736	The Handmaid's Tale	273	3.454212
74755	Girl, Interrupted	251	3.450199
52980	Divine Secrets of the Ya-Ya Sisterhood: A Novel	732	3.448087
184931	The Divine Secrets of the Ya-Ya Sisterhood: A ...	376	3.430851

```
In [70]: popularity_df=num_rating.merge(avg_rating,on='Book-Title')
popularity_df
```

Out[70]:

	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
...
241056	Ã?Â?lpiraten.	2	0.000000
241057	Ã?Â?rger mit Produkt X. Roman.	4	5.250000
241058	Ã?Â?sterlich leben.	1	7.000000
241059	Ã?Â?stlich der Berge.	3	2.666667
241060	Ã?Â?thique en toc	2	4.000000

241061 rows x 3 columns

In [72]: popularity_df=popularity_df[popularity_df['num_ratings']>=250].sort_values('

In [73]: popularity_df

Out[73]:

	Book-Title	num_ratings	avg_rating
80421	Harry Potter and the Order of the Phoenix (Boo...	334	5.571856
191602	The Hobbit : The Enchanting Prelude to The Lor...	281	5.007117
219731	To Kill a Mockingbird	389	4.920308
80440	Harry Potter and the Sorcerer's Stone (Harry P...	571	4.900175
80409	Harry Potter and the Chamber of Secrets (Book 2)	351	4.729345
183564	The Da Vinci Code	883	4.652322
180547	The Catcher in the Rye	403	4.635236
187871	The Five People You Meet in Heaven	427	4.543326
187368	The Fellowship of the Ring (The Lord of the Ri...	257	4.505837
196316	The Lovely Bones: A Novel	1295	4.468726
144158	Prodigal Summer: A Novel	253	4.450593
223125	Tuesdays with Morrie: An Old Man, a Young Man,...	391	4.406650
168711	Stupid White Men ...and Other Sorry Excuses fo...	283	4.356890
204377	The Red Tent (Bestselling Backlist)	723	4.334716
206492	The Secret Life of Bees	615	4.219512
74747	Girl with a Pearl Earring	526	4.218631
129373	Nickel and Dime: On (Not) Getting By in America	267	4.142322
8434	ANGELA'S ASHES	269	4.115242
233360	Where the Heart Is (Oprah's Book Club (Paperba...	585	4.105983
2281	A Child Called \It\": One Child's Courage to S...	262	4.091603
76340	Good in Bed	470	4.040426
64928	Fast Food Nation: The Dark Side of the All- Ame...	321	4.037383
105772	Left Behind: A Novel of the Earth's Last Days ...	297	4.026936
12699	American Gods	302	4.006623
5664	A Prayer for Owen Meany	412	3.781553
212060	The Vampire Lestat (Vampire Chronicles, Book II)	301	3.777409

	Book-Title	num_ratings	avg_rating
233841	White Oleander : A Novel (Oprah's Book Club)	356	3.772472
234730	Wicked: The Life and Times of the Wicked Witch...	326	3.766871
219339	Timeline	407	3.761671
14391	Angels & Demons	586	3.718430
21753	Bel Canto: A Novel	350	3.702857
118232	Me Talk Pretty One Day	340	3.691176
19779	Balzac and the Little Chinese Seamstress : A N...	283	3.657244
134487	One for the Money (Stephanie Plum Novels (Pape...	291	3.656357
38811	Confessions of a Shopaholic (Summer Display Op...	350	3.634286
112927	Lucky : A Memoir	301	3.621262
148031	Red Dragon	365	3.619178
202540	The Poisonwood Bible: A Novel	494	3.609312
199227	The Nanny Diaries: A Novel	474	3.597046
202234	The Pilot's Wife : A Novel	568	3.593310
159293	She's Come Undone (Oprah's Book Club)	468	3.570513
93205	Interview with the Vampire	506	3.535573
818	1st to Die: A Novel	391	3.498721
159292	She's Come Undone (Oprah's Book Club (Paperback))	446	3.495516
223850	Two for the Dough	267	3.464419
60163	Empire Falls	321	3.454829
190736	The Handmaid's Tale	273	3.454212
74755	Girl, Interrupted	251	3.450199
52980	Divine Secrets of the Ya-Ya Sisterhood: A Novel	732	3.448087
184931	The Divine Secrets of the Ya-Ya Sisterhood: A ...	376	3.430851

```
In [75]: popular_df=popularity_df.merge(book,on='Book-Title').drop_duplicates('Book-Title')
```

```
In [76]: popular_df
```


Out[76]:

	Book-Title	Book- Author	Image-URL-M r
0	Harry Potter and the Order of the Phoenix (Boo...	J. K. Rowling	http://images.amazon.com/images/P/043935806X.0...
4	The Hobbit : The Enchanting Prelude to The Lor...	J.R.R. TOLKIEN	http://images.amazon.com/images/P/0345339681.0...
5	To Kill a Mockingbird	Harper Lee	http://images.amazon.com/images/P/0446310786.0...
13	Harry Potter and the Sorcerer's Stone (Harry P...	J. K. Rowling	http://images.amazon.com/images/P/059035342X.0...
15	Harry Potter and the Chamber of Secrets (Book 2)	J. K. Rowling	http://images.amazon.com/images/P/0439064872.0...
18	The Da Vinci Code	Dan Brown	http://images.amazon.com/images/P/0385504209.0...
24	The Catcher in the Rye	J.D. Salinger	http://images.amazon.com/images/P/0316769487.0...
31	The Five People You Meet in Heaven	Mitch Albom	http://images.amazon.com/images/P/0786868716.0...
33	The Fellowship of the Ring (The Lord of the Ri...	J.R.R. TOLKIEN	http://images.amazon.com/images/P/0345339703.0...
42	The Lovely Bones: A Novel	Alice Sebold	http://images.amazon.com/images/P/0316666343.0...
43	Prodigal Summer: A Novel	Barbara Kingsolver	http://images.amazon.com/images/P/0060959037.0...
44	Tuesdays with Morrie: An Old Man, a Young Man,...	MITCH ALBOM	http://images.amazon.com/images/P/0385484518.0...
47	Stupid White Men ...and Other	Michael Moore	http://images.amazon.com/images/P/0060392452.0...

	Book-Title	Book-Author	Image-URL-M r
	Sorry Excuses fo...		
48	The Red Tent (Bestselling Backlist)	Anita Diamant	http://images.amazon.com/images/P/0312195516.0...
49	The Secret Life of Bees	Sue Monk Kidd	http://images.amazon.com/images/P/0142001740.0...
55	Girl with a Pearl Earring	Tracy Chevalier	http://images.amazon.com/images/P/0452282152.0...
56	Nickel and Dimed: On (Not) Getting By in America	Barbara Ehrenreich	http://images.amazon.com/images/P/0805063897.0...
58	ANGELA'S ASHES	Frank McCourt	http://images.amazon.com/images/P/0684874350.0...
60	Where the Heart Is (Oprah's Book Club (Paperba...	Billie Letts	http://images.amazon.com/images/P/0446672211.0...
61	A Child Called \It\": One Child's Courage to S...	Dave Pelzer	http://images.amazon.com/images/P/1558743669.0...
63	Good in Bed	Jennifer Weiner	http://images.amazon.com/images/P/0743418174.0...
65	Fast Food Nation: The Dark Side of the All- Ame...	Eric Schlosser	http://images.amazon.com/images/P/0060938455.0...
66	Left Behind: A Novel of the Earth's Last Days ...	Tim Lahaye	http://images.amazon.com/images/P/0842329129.0...
68	American Gods	Neil Gaiman	http://images.amazon.com/images/P/0380789035.0...
69	A Prayer for Owen Meany	John Irving	http://images.amazon.com/images/P/0345361792.0...
71	The Vampire Lestat (Vampire Chronicles, Book II)	ANNE RICE	http://images.amazon.com/images/P/0345313860.0...

	Book-Title	Book-Author	Image-URL-M r
72	White Oleander : A Novel (Oprah's Book Club)	Janet Fitch	http://images.amazon.com/images/P/0316284955.0...
73	Wicked: The Life and Times of the Wicked Witch...	Gregory Maguire	http://images.amazon.com/images/P/0060987103.0...
74	Timeline	MICHAEL CRICHTON	http://images.amazon.com/images/P/0345417623.0...
81	Angels & Demons	Dan Brown	http://images.amazon.com/images/P/0671027360.0...
86	Bel Canto: A Novel	Ann Patchett	http://images.amazon.com/images/P/0060934417.0...
87	Me Talk Pretty One Day	David Sedaris	http://images.amazon.com/images/P/0316776963.0...
91	Balzac and the Little Chinese Seamstress : A N...	DAI SIJIE	http://images.amazon.com/images/P/0385722206.0...
93	One for the Money (Stephanie Plum Novels (Pape...	Janet Evanovich	http://images.amazon.com/images/P/0061009059.0...
94	Confessions of a Shopaholic (Summer Display Op...	SOPHIE KINSELLA	http://images.amazon.com/images/P/0385335482.0...
95	Lucky : A Memoir	Alice Sebold	http://images.amazon.com/images/P/0316096199.0...
96	Red Dragon	Thomas Harris	http://images.amazon.com/images/P/0440206154.0...
104	The Poisonwood Bible: A Novel	Barbara Kingsolver	http://images.amazon.com/images/P/0060930535.0...
105	The Nanny Diaries: A Novel	Emma McLaughlin	http://images.amazon.com/images/P/0312278586.0...

	Book-Title	Book-Author	Image-URL-M r
107	The Pilot's Wife : A Novel	Anita Shreve	http://images.amazon.com/images/P/0316601950.0...
108	She's Come Undone (Oprah's Book Club)	Wally Lamb	http://images.amazon.com/images/P/0671021001.0...
109	Interview with the Vampire	Anne Rice	http://images.amazon.com/images/P/0345337662.0...
111	1st to Die: A Novel	James Patterson	http://images.amazon.com/images/P/0446610038.0...
113	She's Come Undone (Oprah's Book Club (Paperback))	Wally Lamb	http://images.amazon.com/images/P/0671003755.0...
114	Two for the Dough	Janet Evanovich	http://images.amazon.com/images/P/0671001795.0...
115	Empire Falls	Richard Russo	http://images.amazon.com/images/P/0375726403.0...
118	The Handmaid's Tale	Margaret Atwood	http://images.amazon.com/images/P/0449212602.0...
122	Girl, Interrupted	SUSANNA KAYSEN	http://images.amazon.com/images/P/0679746048.0...
125	Divine Secrets of the Ya-Ya Sisterhood: A Novel	Rebecca Wells	http://images.amazon.com/images/P/0060928336.0...
128	The Divine Secrets of the Ya-Ya Sisterhood: A ...	Rebecca Wells	http://images.amazon.com/images/P/0060502258.0...

In [107... `popular_df.info()`

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 50 entries, 0 to 128
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Book-Title      50 non-null    object
1   Book-Author     50 non-null    object
2   Image-URL-M     50 non-null    object
3   num_ratings     50 non-null    int64
4   avg_rating      50 non-null    float64
dtypes: float64(1), int64(1), object(3)
memory usage: 2.3+ KB

```

```
In [110]: popular_df.columns
```

```
Out[110]: Index(['Book-Title', 'Book-Author', 'Image-URL-M', 'num_ratings',
               'avg_rating'],
              dtype='object')
```

```
In [80]: x=ratings_with_name.groupby('User-ID').count()['Book-Rating']>200
         educated_users=x[x].index
```

```
In [81]: x
```

```
Out[81]: User-ID
2        False
8        False
9        False
10       False
12       False
...
278844   False
278846   False
278849   False
278851   False
278854   False
Name: Book-Rating, Length: 83910, dtype: bool
```

```
In [83]: filtered_ratings=ratings_with_name[ratings_with_name['User-ID'].isin(educated_users)]
```

```
In [84]: filtered_ratings
```

Out[84]:

	User-ID	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher
2	6543	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
5	23768	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
7	28523	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
15	77940	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
16	81977	0	Flesh Tones: A Novel	M. J. Rose	2002	Ballantine Books
...
882841	275970	0	The Theology of the Hammer	Millard Fuller	1994	Smyth & Helwys Publishing
882842	275970	0	The Ordeal of Integration: Progress and Resent...	Orlando Patterson	1998	Civitas Book Publisher
882843	275970	0	Pushcart's Complete Rotten Reviews & Rejec...	Bill Henderson	1998	Pushcart Press
882844	275970	0	There's a Porcupine in My Outhouse: Misadventu...	Mike Tougias	2002	Capital Books (VA)
882845	275970	10	Die Biene.	Sybil GrÃ? Âœfin SchÃ? Â¶nfeldt	1993	Bibliographisches Institut, Mannheim

382064 rows × 6 columns

In [106... `filtered_ratings.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 382064 entries, 2 to 882845
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   User-ID               382064 non-null  int64
1   Book-Rating           382064 non-null  int64
2   Book-Title            382064 non-null  object
3   Book-Author           382064 non-null  object
4   Year-Of-Publication    382064 non-null  object
5   Publisher              382063 non-null  object
dtypes: int64(2), object(4)
memory usage: 20.4+ MB
```

```
In [111]: filtered_ratings.columns
```

```
Out[111]: Index(['User-ID', 'Book-Rating', 'Book-Title', 'Book-Author',
                'Year-Of-Publication', 'Publisher'],
                dtype='object')
```

```
In [85]: y=filtered_ratings.groupby('Book-Title').count()['Book-Rating']>=50
famous_books=y[y].index
```

```
In [86]: y
```

```
Out[86]: Book-Title
A Light in the Storm: The Civil War Diary of Amelia Martin, Fenwick Island, Delaware, 1861 (Dear America)    False
Always Have Popsicles
False
Apple Magic (The Collector's series)
False
Beyond IBM: Leadership Marketing and Finance for the 1990s
False
Clifford Visita El Hospital (Clifford El Gran Perro Colorado)
False

...
Ã?coute ma diffÃ©rence (Le Temps des femmes)
False
Ã?ber das Fernsehen.
False
Ã?ber die Pflicht zum Ungehorsam gegen den Staat.
False
Ã?lpiraten.
False
Ã?stlich der Berge.
False
Name: Book-Rating, Length: 147674, dtype: bool
```

```
In [88]: final_ratings=filtered_ratings[filtered_ratings['Book-Title'].isin(famous_books)]
```

```
In [89]: final_ratings
```

Out[89]:

	User-ID	Book-Rating	Book-Title	Book-Author	Year-Of-Publication	Publisher
63	277427	0	Lightning	Dean R. Koontz	1996	Berkley Publishing Group
69	13552	10	Lightning	Dean R. Koontz	1996	Berkley Publishing Group
70	14521	0	Lightning	Dean R. Koontz	1996	Berkley Publishing Group
72	16795	8	Lightning	Dean R. Koontz	1996	Berkley Publishing Group
74	23768	0	Lightning	Dean R. Koontz	1996	Berkley Publishing Group
...
416891	246655	0	Deck the Halls (Holiday Classics)	Mary Higgins Clark	2002	Pocket
416892	250184	0	Deck the Halls (Holiday Classics)	Mary Higgins Clark	2002	Pocket
416894	256407	0	Deck the Halls (Holiday Classics)	Mary Higgins Clark	2002	Pocket
416895	261829	0	Deck the Halls (Holiday Classics)	Mary Higgins Clark	2002	Pocket
416897	266226	0	Deck the Halls (Holiday Classics)	Mary Higgins Clark	2002	Pocket

26162 rows × 6 columns

In [105... `final_ratings.info()`


```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 26162 entries, 63 to 416897
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   User-ID               26162 non-null  int64
1   Book-Rating           26162 non-null  int64
2   Book-Title            26162 non-null  object
3   Book-Author           26162 non-null  object
4   Year-Of-Publication    26162 non-null  object
5   Publisher              26162 non-null  object
dtypes: int64(2), object(4)
memory usage: 1.4+ MB

```

```
In [112]: final_ratings.columns
```

```
Out[112]: Index(['User-ID', 'Book-Rating', 'Book-Title', 'Book-Author',
                'Year-Of-Publication', 'Publisher'],
               dtype='object')
```

```
In [90]: pt=final_ratings.pivot_table(index='Book-Title',columns='User-ID',values='Book-Rating')
```

```
In [91]: pt
```

Out[91]:

User-ID	254	2276	2766	2977	3363	4385	6251	6543	6575	7158
Book-Title										
1st to Die: A Novel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	9.0	NaN	NaN
4 Blondes	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	NaN	NaN
A Case of Need	NaN	NaN	0.0	NaN	NaN	NaN	0.0	NaN	NaN	NaN
A Child Called \It\": One Child's Courage to Survive"	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
A Heartbreaking Work of Staggering Genius	NaN	NaN	NaN	3.0	NaN	NaN	4.0	0.0	7.0	NaN
...
Year of Wonders	NaN	NaN	NaN	7.0	NaN	NaN	NaN	0.0	0.0	NaN
You Belong To Me	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0
Zen and the Art of Motorcycle Maintenance: An Inquiry into Values	NaN	NaN	NaN	NaN	0.0	NaN	0.0	NaN	NaN	NaN
Zoya	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
\O\" Is for Outlaw"	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

342 rows × 665 columns

In [92]: `pt.fillna(0,inplace=True)`

In [93]: `pt`

Out[93]:

User-ID	254	2276	2766	2977	3363	4385	6251	6543	6575	7158
Book-Title										
1st to Die: A Novel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	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
A Case of Need	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A Child Called It: One Child's Courage to Survive"	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A Heartbreaking Work of Staggering Genius	0.0	0.0	0.0	3.0	0.0	0.0	4.0	0.0	7.0	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
You Belong To Me	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
Zoya	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

342 rows × 665 columns

```
In [95]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [96]: similarity_score=cosine_similarity(pt)
similarity_score
```

```
Out[96]: array([[1.          , 0.          , 0.02085258, ..., 0.04702034, 0.14187761,
                0.21732542],
               [0.          , 1.          , 0.14526454, ..., 0.          , 0.          ,
                0.          ],
               [0.02085258, 0.14526454, 1.          , ..., 0.          , 0.          ,
                0.          ],
               ...,
               [0.04702034, 0.          , 0.          , ..., 1.          , 0.07499688,
                0.05053319],
               [0.14187761, 0.          , 0.          , ..., 0.07499688, 1.          ,
                0.1757118 ],
               [0.21732542, 0.          , 0.          , ..., 0.05053319, 0.1757118 ,
                1.          ]])
```

```
In [99]: print(pt.index.tolist()) # This will give me a list of all book titles
```

['1st to Die: A Novel', '4 Blondes', 'A Case of Need', 'A Child Called \\It \\": One Child\\'s Courage to Survive"', 'A Heartbreaking Work of Staggering Genius', 'A Is for Alibi (Kinsey Millhone Mysteries (Paperback))', 'A Painted House', 'A Prayer for Owen Meany', 'A Thin Dark Line (Mysteries & Horror)', 'A Thousand Acres (Ballantine Reader's Circle)', 'A Time to Kill', 'A Virtuous Woman (Oprah's Book Club (Paperback))', 'A Walk to Remember', 'A Wrinkle In Time', 'A Wrinkle in Time', 'ANGELA'S ASHES', 'Airframe', 'All Around the Town', 'All I Really Need to Know', 'All That Remains (Kay Scarpetta Mysteries (Paperback))', 'All the Pretty Horses (The Border Trilogy, Vol 1)', 'Along Came a Spider (Alex Cross Novels)', 'American Gods', 'Angela's Ashes (MMP) : A Memoir', 'Angela's Ashes: A Memoir', 'Angels & Demons', 'Animal Dreams', 'Anne Frank: The Diary of a Young Girl', 'Anne of Avonlea (Anne of Green Gables Novels (Paperback))', 'Anne of Green Gables (Anne of Green Gables Novels (Paperback))', 'Artemis Fowl (Artemis Fowl, Book 1)', 'Ashes to Ashes', 'At Home in Mitford (The Mitford Years)', 'Atonement : A Novel', 'B Is for Burglar (Kinsey Millhone Mysteries (Paperback))', 'Back Roads', 'Balzac and the Little Chinese Seamstress : A Novel', 'Before I Say Good-Bye', 'Bel Canto: A Novel', 'Betrayal in Death', 'Big Stone Gap: A Novel (Ballantine Reader's Circle)', 'Black and Blue', 'Body of Evidence (Kay Scarpetta Mysteries (Paperback))', 'Breath, Eyes, Memory', 'Breathing Lessons', 'Bridget Jones: The Edge of Reason', 'C Is for Corpse (Kinsey Millhone Mysteries (Paperback))', 'CAT'S EYE', 'Call of the Wild', 'Carolina Moon', 'Cat & Mouse (Alex Cross Novels)', 'Cause of Death', 'Charlotte's Web (Trophy Novel)', 'Chicken Soup for the Soul (Chicken Soup for the Soul)', 'Chicken Soup for the Woman's Soul (Chicken Soup for the Soul Series (Paper))', 'Circle of Friends', 'Clear and Present Danger (Jack Ryan Novels)', 'Cold Mountain : A Novel', 'Confessions of a Shopaholic (Summer Display Opportunity)', 'Confessions of an Ugly Stepsister : A Novel', 'Congo', 'Cruel & Unusual (Kay Scarpetta Mysteries (Paperback))', 'D Is for Deadbeat (Kinsey Millhone Mysteries (Paperback))', 'Daisy Fay and the Miracle Man', 'Dance upon the Air (Three Sisters Island Trilogy)', 'Dances With Wolves', 'Dark Paradise', 'Dawn (Cutler)', 'Debt of Honor (Jack Ryan Novels)', 'Deception Point', 'Deck the Halls (Holiday Classics)', 'Degree of Guilt', 'Disclosure', 'Divine Secrets of the Ya-Ya Sisterhood: A Novel', 'Dolores Claiborne', 'Don't Sweat the Small Stuff and It's All Small Stuff : Simple Ways to Keep the Little Things from Taking Over Your Life (Don't Sweat the Small Stuff Series)', 'Dragonfly in Amber', 'Drowning Ruth (Oprah's Book Club)', 'Dying for Chocolate (Culinary Mysteries (Paperback))', 'E Is for Evidence: A Kinsey Millhone Mystery (Kinsey Millhone Mysteries (Paperback))', 'Ellen Foster', 'Empire Falls', 'F Is for Fugitive (Kinsey Millhone Mysteries (Paperback))', 'FORREST GUMP (Movie Tie in)', 'Face the Fire (Three Sisters Island Trilogy)', 'Faking It', 'Fall On Your Knees (Oprah #45)', 'False Memory', 'Fast Food Nation: The Dark Side of the All-American Meal', 'Fear Nothing', 'Final Target', 'Fine Things', 'Five Quarters of the Orange', 'Forever... : A Novel of Good and Evil, Love and Hope', 'Four To Score (A Stephanie Plum Novel)', 'Fried Green Tomatoes at the Whistle Stop Cafe', 'From Potter's Field', 'Full House (Janet Evanovich's Full Series)', 'Full Tilt (Janet Evanovich's Full Series)', 'G Is for Gumshoe (Kinsey Millhone Mysteries (Paperback))', 'Girl with a Pearl Earring', 'Girl, Interrupted', 'Good in Bed', 'Guilty as Sin', 'H Is for Homicide (Kinsey Millhone Mysteries (Paperback))', 'Hannibal', 'Hard Eight : A Stephanie Plum Novel (A Stephanie Plum Novel)', 'Harry Potter and the Chamber of Secrets (Book 2)', 'Harry Potter and the Goblet of Fire (Book 4)', 'Harry Potter and the Order of the Phoenix (Book 5)', 'Harry Potter and the Prisoner of Azkaban (Book 3)', 'Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))', 'Heart of the Sea (Irish Trilogy)', 'Hearts In Atlantis', 'Heaven and Earth (Three Sisters Island Trilogy)', 'Here on Earth', 'Hide & S

eeek', 'Hideaway', 'High Five (A Stephanie Plum Novel)', 'Hot Six : A Stephanie Plum Novel (A Stephanie Plum Novel)', 'House of Sand and Fog', 'How to Make an American Quilt', 'I Is for Innocent', 'I Know This Much Is True', 'I Know Why the Caged Bird Sings', 'I'll Be Seeing You', 'Icy Sparks', 'Inner Harbor (Quinn Brothers (Paperback))', 'Interview with the Vampire', 'Into the Wild', 'Invasion', 'Isle of Dogs', 'J Is for Judgment', 'Jack & Jill (Alex Cross Novels)', 'Jemima J: A Novel About Ugly Ducklings and Swans', 'Jewel', 'Jewel (Oprah's Book Club)', 'Jewels of the Sun (Irish Trilogy)', 'Jurassic Park', 'K Is for Killer (Kinsey Millhone Mysteries (Paperback))', 'Key of Knowledge (Key Trilogy (Paperback))', 'Key of Light (Key Trilogy (Paperback))', 'Key of Valor (Roberts, Nora. Key Trilogy, 3.)', 'Kiss the Girls', 'L Is for Lawless', 'Lamb : The Gospel According to Biff, Christ's Childhood Pal', 'Left Behind: A Novel of the Earth's Last Days (Left Behind No. 1)', 'Let Me Call You Sweetheart', 'Lightning', 'Like Water for Chocolate: A Novel in Monthly Installments With Recipes, Romances and Home Remedies', 'Little Altars Everywhere: A Novel', 'Lord of the Flies', 'Love in the Time of Cholera (Penguin Great Books of the 20th Century)', 'Loves Music, Loves to Dance', 'Lucky : A Memoir', 'M Is for Malice', 'Macgregor Grooms (Macgregors)', 'Me Talk Pretty One Day', 'Midnight in the Garden of Good and Evil: A Savannah Story', 'Midwives: A Novel', 'Moonlight Becomes You', 'Morality for Beautiful Girls (No.1 Ladies Detective Agency)', 'Mystic River', 'N Is for Moose', 'Neverwhere', 'Nickel and Dimed: On (Not) Getting By in America', 'Nicolae: The Rise of Antichrist (Left Behind No. 3)', 'Night', 'Night Sins', 'Night Whispers', 'On the Street Where You Live', 'One Door Away from Heaven', 'One for the Money (A Stephanie Plum Novel)', 'One for the Money (Stephanie Plum Novels (Paperback))', 'P Is for Peril', 'Patriot Games (Jack Ryan Novels)', 'Pay It Forward', 'Plainsong (Vintage Contemporaries)', 'Point of Origin', 'Politically Correct Bedtime Stories: Modern Tales for Our Life and Times', 'Postmortem', 'Presumed Innocent', 'Pretend You Don't See Her', 'Purity in Death', 'Red Dragon', 'Red Storm Rising', 'Remember Me', 'Rising Sun', 'River's End', 'River, Cross My Heart', 'Roses Are Red (Alex Cross Novels)', 'SHIPPING NEWS', 'Sanctuary', 'Saving Faith', 'Scarlett : The Sequel to Margaret Mitchell's \\Gone With the Wind\\'", 'Schindler's List', 'Sea Swept (Quinn Brothers (Paperback))', 'Seabiscuit', 'Seabiscuit: An American Legend', 'Secret History', 'She's Come Undone (Oprah's Book Club (Paperback))', 'She's Come Undone (Oprah's Book Club)', 'Shell Seekers', 'Shopaholic Takes Manhattan (Summer Display Opportunity)', 'Shopaholic Ties the Knot', 'Silence of the Lambs', 'Skipping Christmas', 'Songs in Ordinary Time (Oprah's Book Club (Paperback))', 'Sphere', 'Still Waters', 'Stupid White Men ...and Other Sorry Excuses for the State of the Nation!', 'Summer Pleasures', 'Tara Road', 'Tears of the Giraffe (No.1 Ladies Detective Agency)', 'Tears of the Moon (Irish Trilogy)', 'Tell Me Lies (Tell Me Lies)', 'Tell No One', 'The Alienist', 'The Angel of Darkness', 'The Bachelor (Warner Books Contemporary Romance)', 'The Bad Beginning (A Series of Unfortunate Events, Book 1)', 'The Bad Place', 'The Beach House', 'The Bean Trees', 'The Body Farm', 'The Bonfire of the Vanities', 'The Book of Ruth (Oprah's Book Club (Paperback))', 'The Bourne Identity', 'The Bridges of Madison County', 'The Burden of Proof', 'The Cardinal of the Kremlin (Jack Ryan Novels)', 'The Catcher in the Rye', 'The Celestine Prophecy (Celestine Prophecy)', 'The Clan of the Cave Bear (Earth's Children (Paperback))', 'The Color of Water: A Black Man's Tribute to His White Mother', 'The Copper Beech', 'The Da Vinci Code', 'The Dark Half', 'The Divine Secrets of the Ya-Ya Sisterhood: A Novel', 'The Dogs of Babel (Today Show Book Club #12)', 'The Door to December', 'The English Patient', 'The Fellowship of the Ring (The Lord of the Rings, Part 1)', 'The Five People You Meet in Heaven', 'The Girl Who Loved Tom Gordon', 'The Girls' Guide to Hunting and Fishing', 'The Giver (21st Century Reference)', 'The Glass Lake', 'The God of S

mall Things', 'The Handmaid's Tale', 'The Hobbit : The Enchanting Prelude to The Lord of the Rings', 'The Horse Whisperer', 'The Hot Zone', 'The Hours: A Novel', 'The House of the Spirits', 'The Idiot Girls' Action Adventure Club', 'The Joy Luck Club', 'The King of Torts', 'The Last Precinct', 'The Last Time They Met : A Novel', 'The Loop', 'The Lost World', 'The Lottery Winner : Alvirah And Willy Stories', 'The Lovely Bones: A Novel', 'The Mists of Avalon', 'The Mulberry Tree', 'The Mummy or Ramses the Damned', 'The Murder Book', 'The Nanny Diaries: A Novel', 'The No. 1 Ladies' Detective Agency (Today Show Book Club #8)', 'The Notebook', 'The Pelican Brief', 'The Pillars of the Earth', 'The Pilot's Wife : A Novel', 'The Plains of Passage (Earth's Children (Paperback))', 'The Poisonwood Bible: A Novel', 'The Prince of Tides', 'The Queen of the Damned (Vampire Chronicles (Paperback))', 'The Rainmaker', 'The Rapture of Canaan', 'The Red Tent (Bestselling Backlist)', 'The Reef', 'The Rescue', 'The Return Journey', 'The Return of the King (The Lord of the Rings, Part 3)', 'The Robber Bride', 'The Runaway Jury', 'The Saving Graces: A Novel', 'The Secret Life of Bees', 'The Shipping News : A Novel', 'The Simple Truth', 'The Stone Diaries', 'The Street Lawyer', 'The Sum of All Fears (Jack Ryan Novels)', 'The Summerhouse', 'The Summons', 'The Surgeon', 'The Sweet Potato Queens' Book of Love', 'The Tale of the Body Thief (Vampire Chronicles (Paperback))', 'The Tao of Pooh', 'The Testament', 'The Thorn Birds', 'The Tommyknockers', 'The Ugly Duckling', 'The Vampire Lestat (Vampire Chronicles, Book II)', 'The Weight of Water : A Novel Tag: Author of Resistance and Strange Fits of Passion', 'The Winner', 'The Witching Hour (Lives of the Mayfair Witches)', 'Three Junes', 'Three To Get Deadly : A Stephanie Plum Novel (A Stephanie Plum Novel)', 'Timeline', 'To Kill a Mockingbird', 'To the Nines: A Stephanie Plum Novel', 'Tom Clancy's Op-Center (Tom Clancy's Op Center (Paperback))', 'Touching Evil', 'Tribulation Force: The Continuing Drama of Those Left Behind (Left Behind No. 2)', 'Truly, Madly Manhattan', 'Tuesdays with Morrie: An Old Man, a Young Man, and Life's Greatest Lesson', 'Two for the Dough', 'Under the Tuscan Sun', 'Vinegar Hill (Oprah's Book Club (Paperback))', 'Violets Are Blue', 'We Were the Mulvaney', 'We'll Meet Again', 'Welcome to Temptation', 'Welcome to the World, Baby Girl!', 'What Looks Like Crazy On An Ordinary Day', 'What to Expect When You're Expecting (Revised Edition)', 'Where the Heart Is (Oprah's Book Club (Paperback))', 'Where the Red Fern Grows', 'While I Was Gone', 'While My Pretty One Sleeps', 'Whirlwind (Tyler, Book 1)', 'White Oleander : A Novel (Oprah's Book Club)', 'Wicked: The Life and Times of the Wicked Witch of the West', 'Wild Animus', 'Without Remorse', 'Year of Wonders', 'You Belong To Me', 'Zen and the Art of Motorcycle Maintenance: An Inquiry into Values', 'Zoya', '\\0\\' Is for Outlaw"]

```
In [97]: def recommend(book_name):
        index=np.where(pt.index==book_name)[0][0]
        similar_items=sorted(list(enumerate(similarity_score[index])),key=lambda
        for i in similar_items:
            book_title = pt.index[i[0]]
            print(book_title)
```

```
In [100]: def recommend(book_name):
        try:
            index = pt.index.get_loc(book_name)
        except KeyError:
            print(f"Book '{book_name}' not found in the index.")
            return

        similar_items = sorted(list(enumerate(similarity_score[index])), key=lambda
```

```

    for i in similar_items:
        book_title = pt.index[i[0]]
        print(book_title)

recommend('1984')

```

Book '1984' not found in the index.

```

In [102]: def recommend(book_name):
    if book_name not in pt.index:
        print(f"Book '{book_name}' not found in the index.")
        return

    index = np.where(pt.index == book_name)[0][0]
    similar_items = sorted(list(enumerate(similarity_score[index])), key=lambda

    for i in similar_items:
        book_title = pt.index[i[0]]
        print(book_title)

recommend('The Bad Place')

```

Hideaway
The Tommyknockers
Lightning
The Dark Half
We'll Meet Again

similarity_score: This variable is expected to be a 2D array (like a matrix) that holds the similarity scores between books. Each row corresponds to a book, and the values in that row represent the similarity to other books

The recommend function suggests similar books based on a given title. It checks if the book exists, retrieves its index, calculates similarity scores from a matrix, and prints the titles of the top 5 most similar books, excluding the original book itself.

```

In [127]: recommend('Fried Green Tomatoes at the Whistle Stop Cafe')

```

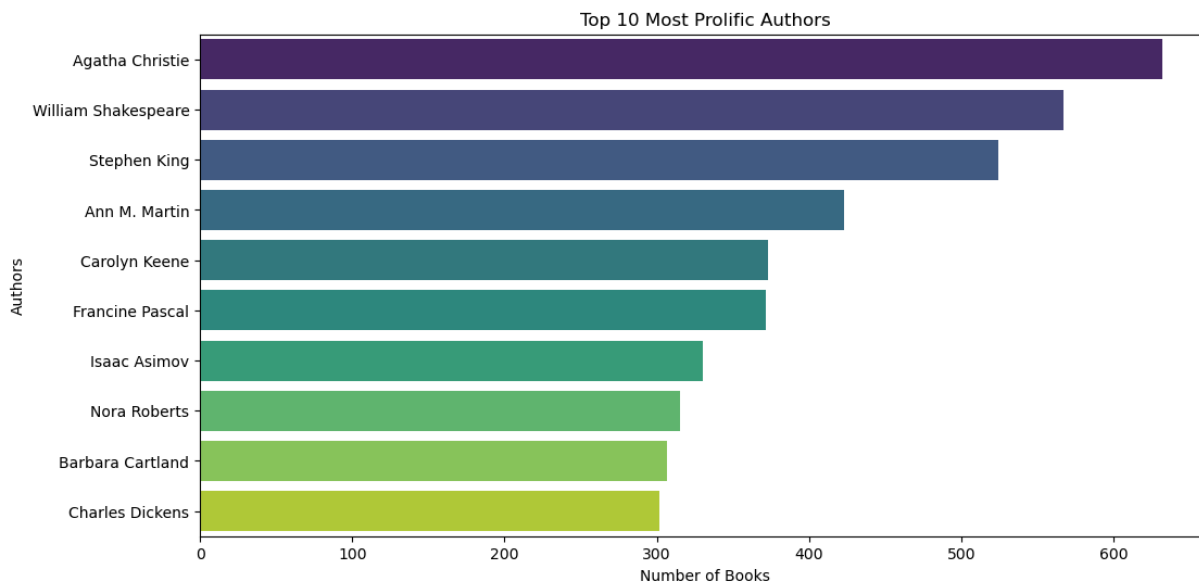

The Red Tent (Bestselling Backlist)
Empire Falls
Songs in Ordinary Time (Oprah's Book Club (Paperback))
The Stone Diaries
Ellen Foster

```
In [126... recommend('Three Junes')
```

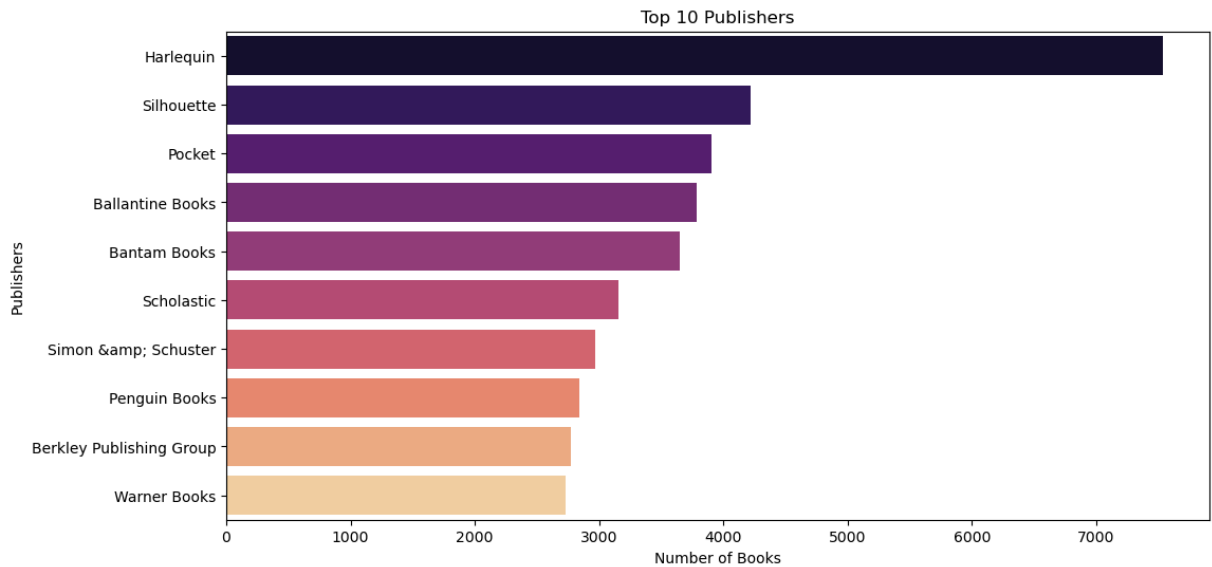
Plainsong (Vintage Contemporaries)
Midnight in the Garden of Good and Evil: A Savannah Story
Seabiscuit: An American Legend
Cold Mountain : A Novel
Morality for Beautiful Girls (No.1 Ladies Detective Agency)

Visualization 1: Most Prolific Authors

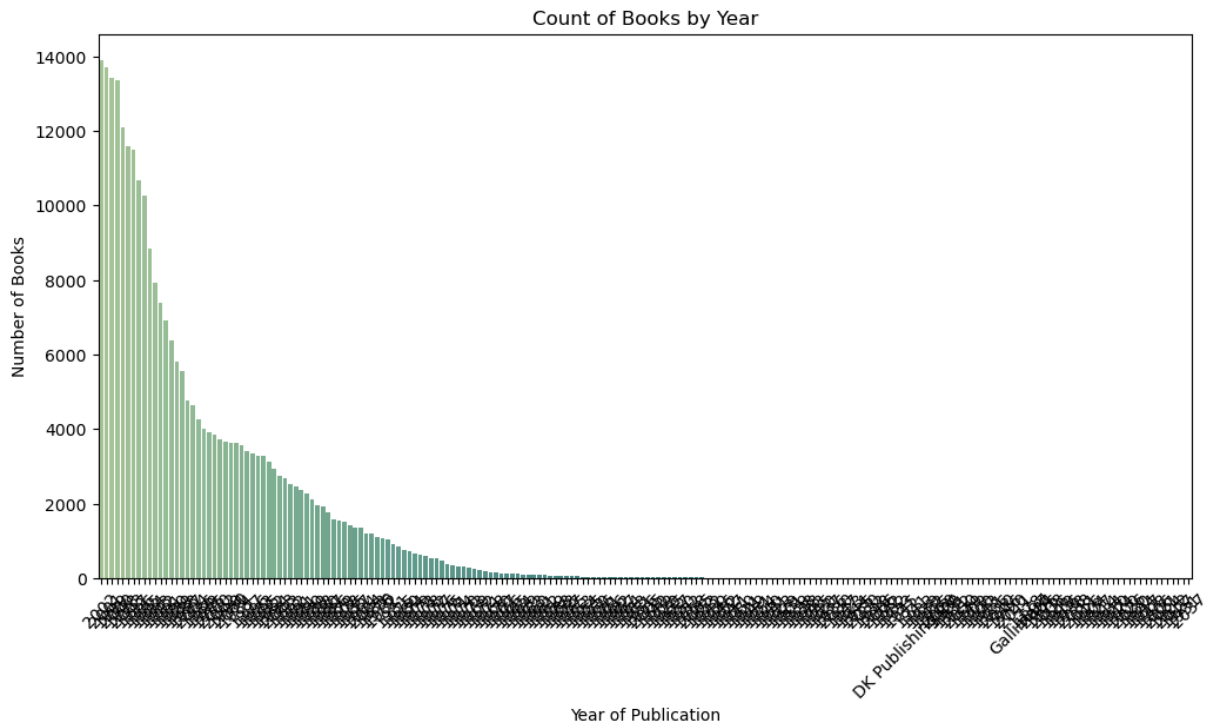
```
In [114... plt.figure(figsize=(12, 6))
top_authors = book['Book-Author'].value_counts().head(10)
sns.barplot(x=top_authors.values, y=top_authors.index, palette='viridis')
plt.title('Top 10 Most Prolific Authors')
plt.xlabel('Number of Books')
plt.ylabel('Authors')
plt.show()
```



```
In [116... # Top Publishers
plt.figure(figsize=(12, 6))
top_publishers = book['Publisher'].value_counts().head(10)
sns.barplot(x=top_publishers.values, y=top_publishers.index, palette='magma')
plt.title('Top 10 Publishers')
plt.xlabel('Number of Books')
plt.ylabel('Publishers')
plt.show()
```



```
In [118... # Count of Books by Year
plt.figure(figsize=(12, 6))
sns.countplot(data=book, x='Year-Of-Publication', palette='crest', order=book
plt.title('Count of Books by Year')
plt.xlabel('Year of Publication')
plt.ylabel('Number of Books')
plt.xticks(rotation=45)
plt.show()
```



In []:

In []:

```
In [119... import pickle
pickle.dump(popular,open('popular.pkl','wb'))
```

```
In [ ]:
```

```
In [ ]:
```