

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
books_df = pd.read_csv("books.csv",on_bad_lines='skip')
```

In [3]:

```
books_df.shape
```

Out[3]:

```
(11123, 12)
```

In [4]:

```
books_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11123 entries, 0 to 11122
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   bookID                11123 non-null  int64  
 1   title                 11123 non-null  object  
 2   authors               11123 non-null  object  
 3   average_rating        11123 non-null  float64  
 4   isbn                  11123 non-null  object  
 5   isbn13                11123 non-null  int64  
 6   language_code         11123 non-null  object  
 7   num_pages             11123 non-null  int64  
 8   ratings_count         11123 non-null  int64  
 9   text_reviews_count    11123 non-null  int64  
10   publication_date       11123 non-null  object  
11   publisher              11123 non-null  object  
dtypes: float64(1), int64(5), object(6)
memory usage: 1.0+ MB
```

In [5]:

```
books_df.columns
```

Out[5]:

```
Index(['bookID', 'title', 'authors', 'average_rating', 'isbn', 'isbn13',
       'language_code', ' num_pages', 'ratings_count', 'text_reviews_coun
t',
       'publication_date', 'publisher'],
      dtype='object')
```

In [6]:

```
books_df.describe(include="all")
```

Out[6]:

	bookID	title	authors	average_rating	isbn	isbn13	language_cod
count	11123.000000	11123	11123	11123.000000	11123	1.112300e+04	1112
unique	NaN	10348	6639	NaN	11123	NaN	2
top	NaN	The Iliad	Stephen King	NaN	0439785960	NaN	en
freq	NaN	9	40	NaN	1	NaN	890
mean	21310.856963	NaN	NaN	3.934075	NaN	9.759880e+12	NaN
std	13094.727252	NaN	NaN	0.350485	NaN	4.429758e+11	NaN
min	1.000000	NaN	NaN	0.000000	NaN	8.987060e+09	NaN
25%	10277.500000	NaN	NaN	3.770000	NaN	9.780345e+12	NaN
50%	20287.000000	NaN	NaN	3.960000	NaN	9.780582e+12	NaN
75%	32104.500000	NaN	NaN	4.140000	NaN	9.780872e+12	NaN
max	45641.000000	NaN	NaN	5.000000	NaN	9.790008e+12	NaN

In [7]:

```
books_df.rename(columns={' num_pages ':'num_page'}, inplace=True)
```

In [8]:

```
rating_count_mean = books_df["ratings_count"].mean()
rating_count_mean
```

Out[8]:

17942.848062573048

In [9]:

```
avg_rating_mean = books_df["average_rating"].mean()
avg_rating_mean
```

Out[9]:

3.934075339386829

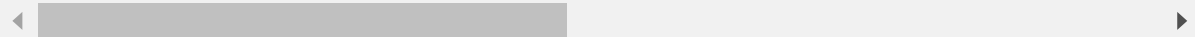
In [10]:

```
# high rated books
books_df.sort_values(by=['average_rating'],ascending=False)
```

Out[10]:

	bookID		title	authors	average_rating	isbn
624	2034	Acharenses/Equites/Nubes/Vespae/P...	Comoediae 1: Aristophanes/F.W. Hall/W.M. Geldart		5.0	0198145047
9893	39829	His Princess Devotional: A Royal Encounter Wit...	Sheri Rose Shepherd		5.0	1590529626
4788	17224	The Diamond Color Meditation: Color Pathway to...	John Diamond		5.0	1890995525
9324	36853	Tyrannosaurus Wrecks (Stanley #1)	Laura Driscoll/Alisa Klayman-Grodsky/Eric ...		5.0	0786845031
9720	38804	The Irish Anatomist: A Study of Flann O'Brien	Keith Donohue		5.0	1930901356
...	...		...	...	...	...
6877	25903	I'll Be Home Before Midnight and I Won't Get P...	Anthony E. Wolf		0.0	0394755669
3492	12712	Brodie's notes on Aldous Huxley's brave new world	Graham Handley		0.0	0333581296
2533	9338	Insights: Talks On The Nature Of Existence	Frederick P. Lenz		0.0	1932206086
2532	9337	How To Meditate: An Anthology Of Talks On Medi...	Frederick P. Lenz		0.0	1932206108
10497	42758	The Fantastic Vampire: Studies in the Children...	James Craig Holte		0.0	0313309337

11123 rows × 12 columns



In [11]:

```
# popular books
books_df["w_rating"] = books_df["average_rating"]*books_df["ratings_count"]
```

In [12]:

```
books_df.sort_values(by=['w_rating'],ascending=False)
```

Out[12]:

	bookID	title	authors	average_rating	isbn	isbn13	lang
10336	41865	Twilight (Twilight #1)	Stephenie Meyer	3.59	0316015849	9780316015844	
1697	5907	The Hobbit or There and Back Again	J.R.R. Tolkien	4.27	0618260307	9780618260300	
3	5	Harry Potter and the Prisoner of Azkaban (Harr...	J.K. Rowling/Mary GrandPré	4.56	043965548X	9780439655484	
4415	15881	Harry Potter and the Chamber of Secrets (Harry...	J.K. Rowling/Mary GrandPré	4.42	0439064864	9780439064866	
1	2	Harry Potter and the Order of the Phoenix (Har...	J.K. Rowling/Mary GrandPré	4.49	0439358078	9780439358071	
...	...	...	...	...	...	...	
264	797	Lonely Planet Londres	Lonely Planet/Sarah Johnstone/Tom Masters	4.03	8408064762	9788408064763	
265	799	Out to Eat London 2002 (Lonely Planet Out to Eat)	Lonely Planet/Mark Honan	0.00	1740592050	9781740592055	
6999	26525	Leadership Challenge	James M. Kouzes	4.05	0787983500	9780787983505	
10862	44433	Courageous Faith Through the Year	Bill Hybels/Keri Wyatt Kent	3.50	0830832947	9780830832941	
5779	21703	The Best American Mystery Stories 2002	James Ellroy/Otto Penzler/John Biguenet/Michae...	3.58	0618258078	9780618258079	

11123 rows × 13 columns



In [13]:

```
books_df[books_df['w_rating']==max(books_df["w_rating"])]
```

Out[13]:

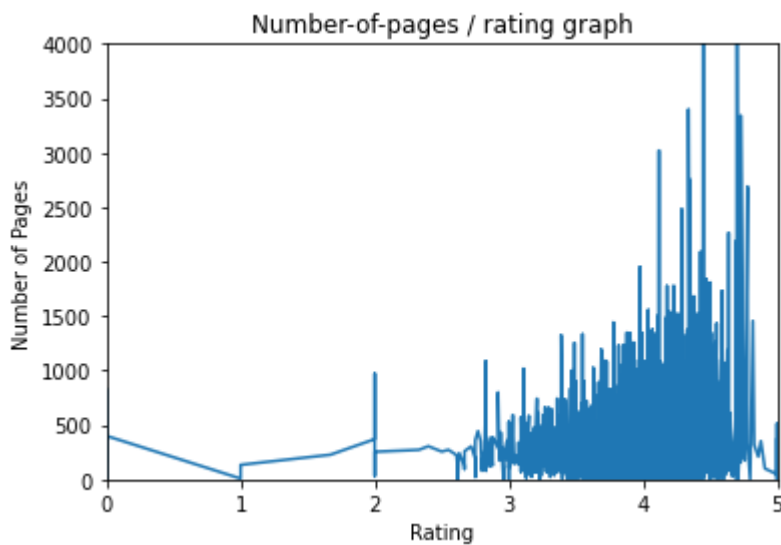
	bookID	title	authors	average_rating	isbn	isbn13	language_code
10336	41865	Twilight (Twilight #1)	Stephenie Meyer	3.59	0316015849	9780316015844	eng

In [14]:

```
avg_sort =books_df.sort_values(by=["average_rating"],ascending=False)
```

In [15]:

```
plt.plot(avg_sort["average_rating"],avg_sort["num_page"])
plt.xlim(0,5)
plt.xlabel("Rating")
plt.ylabel("Number of Pages")
plt.ylim(0,4000)
plt.title("Number-of-pages / rating graph")
plt.show()
```



In [16]:

```
books_df.head(2)
```

Out[16]:

	bookID	title	authors	average_rating	isbn	isbn13	language_code
0	1	Harry Potter and the Half-Blood Prince (Harry ...	J.K. Rowling/Mary GrandPré	4.57	0439785960	9780439785969	eng
1	2	Harry Potter and the Order of the Phoenix (Har...	J.K. Rowling/Mary GrandPré	4.49	0439358078	9780439358071	eng

In [ ]:

In [17]:

```
for i in range(len(books_df["publication_date"])):
    books_df["publication_date"][i]=books_df["publication_date"][i][-4:]
```

C:\Users\RAHUL SUTHAR\AppData\Local\Temp\ipykernel\_14388\4040627211.py:2: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))

```
books_df["publication_date"][i]=books_df["publication_date"][i][-4:]
```

In [18]:

books\_df

Out[18]:

	bookID	title	authors	average_rating	isbn	isbn13	lan
0	1	Harry Potter and the Half-Blood Prince (Harry ...	J.K. Rowling/Mary GrandPré	4.57	0439785960	9780439785969	
1	2	Harry Potter and the Order of the Phoenix (Har...	J.K. Rowling/Mary GrandPré	4.49	0439358078	9780439358071	
2	4	Harry Potter and the Chamber of Secrets (Harry...	J.K. Rowling	4.42	0439554896	9780439554893	
3	5	Harry Potter and the Prisoner of Azkaban (Harr...	J.K. Rowling/Mary GrandPré	4.56	043965548X	9780439655484	
4	8	Harry Potter Boxed Set Books 1-5 (Harry Potte...	J.K. Rowling/Mary GrandPré	4.78	0439682584	9780439682589	
...	...	...	...	...	...	...	...
11118	45631	Expelled from Eden: A William T. Vollmann Reader	William T. Vollmann/Larry McCaffery/Michael He...	4.06	1560254416	9781560254416	
11119	45633	You Bright and Risen Angels	William T. Vollmann	4.08	0140110879	9780140110876	
11120	45634	The Ice-Shirt (Seven Dreams #1)	William T. Vollmann	3.96	0140131965	9780140131963	
11121	45639	Poor People	William T. Vollmann	3.72	0060878827	9780060878825	

	bookID	title	authors	average_rating	isbn	isbn13	lan
11122	45641	Las aventuras de Tom Sawyer	Mark Twain	3.91	8497646983	9788497646987	

11123 rows × 13 columns

In [19]:

```
new_df3 = books_df[books_df["publication_date"]=="1960"]
```

In [20]:

```
new_df3.shape
```

Out[20]:

(6, 13)

In [21]:

```
new_df3
```

Out[21]:

	bookID	title	authors	average_rating	isbn	isbn13	lan
963	3258	Greek Tragedies Volume 2	David Grene/Richmond Lattimore/Aeschylus/Sopho...	4.29	0226307751	9780226307751	
3354	12250	Greek Lyrics	Richmond Lattimore	3.98	0226469441	9780226469441	
3589	13018	Cliffsnotes on Shakespeare's Henry IV Part 1	James K. Lowers/William Shakespeare/CliffsNotes	3.88	0822000237	9780822000237	
7240	27800	Faust Part Two	Johann Wolfgang von Goethe/Philip Wayne	3.73	0140440933	9780140440933	
9809	39361	Paws Before Dying (A Dog Lover's Mystery #4)	Susan Conant	3.90	0425144305	9780425144305	
10459	42594	CliffsNotes on Shakespeare's Twelfth Night	James Lamar Roberts/CliffsNotes/William Shakes...	4.16	0822000946	9780822000946	

In [ ]:

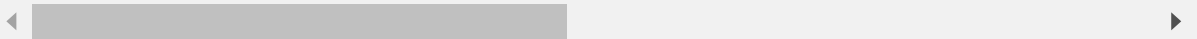


In [22]:

```
#most popular book of the 60s?  
new_df3.sort_values(by=["w_rating"],ascending=False)
```

Out[22]:

	bookID	title	authors	average_rating	isbn	is
9809	39361	Paws Before Dying (A Dog Lover's Mystery #4)	Susan Conant	3.90	0425144305	97804251.
963	3258	Greek Tragedies Volume 2	David Grene/Richmond Lattimore/Aeschylus/Sopho...	4.29	0226307751	97802263
3354	12250	Greek Lyrics	Richmond Lattimore	3.98	0226469441	97802264
7240	27800	Faust Part Two	Johann Wolfgang von Goethe/Philip Wayne	3.73	0140440933	97801404.
3589	13018	Cliffsnotes on Shakespeare's Henry IV Part 1	James K. Lowers/William Shakespeare/CliffsNotes	3.88	0822000237	97808220
10459	42594	CliffsNotes on Shakespeare's Twelfth Night	James Lamar Roberts/CliffsNotes/William Shakes...	4.16	0822000946	97808220



In [23]:

```
books_df.sort_values(by=["authors"])
```

Out[23]:

	bookID	title	authors	average_rating	isbn	isbn13	lar
8037	30855	A Woman in Jerusalem	A.B. Yehoshua/Hillel Halkin	3.43	0151012261	9780151012268	
7169	27503	The Liberated Bride	A.B. Yehoshua/Hillel Halkin	3.77	0156030160	9780156030168	
4952	17883	Notes from Underground & A Confession (Everyma...	A.D.P. Briggs/Leo Tolstoy/Fyodor Dostoyevsky	3.76	0460874489	9780460874489	
10110	40880	Jack Vance: Critical Appreciations and a Bibli...	A.E. Cunningham/Harlan Ellison/Charles F. Mill...	4.15	0712311025	9780712311021	
7300	28116	The Know-It-All: One Man's Humble Quest to Bec...	A.J. Jacobs	3.76	0743250621	9780743250627	
...	...	...	...	...	...	...	
7383	28419	La Débâcle	Émile Zola/Robert Lethbridge/Elinor Dorday	4.01	0192822896	9780192822895	
7379	28413	The Ladies' Paradise	Émile Zola/Robin Buss/Brian Nelson	3.99	0192836021	9780192836021	
7377	28409	The Masterpiece	Émile Zola/Roger Pearson	3.96	0192839632	9780192839633	
7375	28407	Germinal	Émile Zola/Roger Pearson	4.13	0140447423	9780140447422	
1246	4314	Monsieur Ibrahim et les fleurs du Coran	Éric-Emmanuel Schmitt	3.82	2226126260	9782226126269	

11123 rows × 13 columns

In [24]:

```
df =books_df.groupby("authors",as_index=False)['num_page'].sum()
```

In [25]:

```
df =pd.DataFrame(df)
df
```

Out[25]:

	authors	num_page
0	A.B. Yehoshua/Hillel Halkin	832
1	A.D.P. Briggs/Leo Tolstoy/Fyodor Dostoyevsky	256
2	A.E. Cunningham/Harlan Ellison/Charles F. Mill...	232
3	A.J. Jacobs	389
4	A.M. Homes	256
...	...	...
6634	Émile Zola/Henri Mitterand	462
6635	Émile Zola/Robert Lethbridge/Elinor Dorday	592
6636	Émile Zola/Robin Buss/Brian Nelson	438
6637	Émile Zola/Roger Pearson	1056
6638	Éric-Emmanuel Schmitt	85

6639 rows × 2 columns

In [26]:

```
# most page written by stephen king
df_sort =df.sort_values(by=['num_page'],ascending=False)
df_sort
```

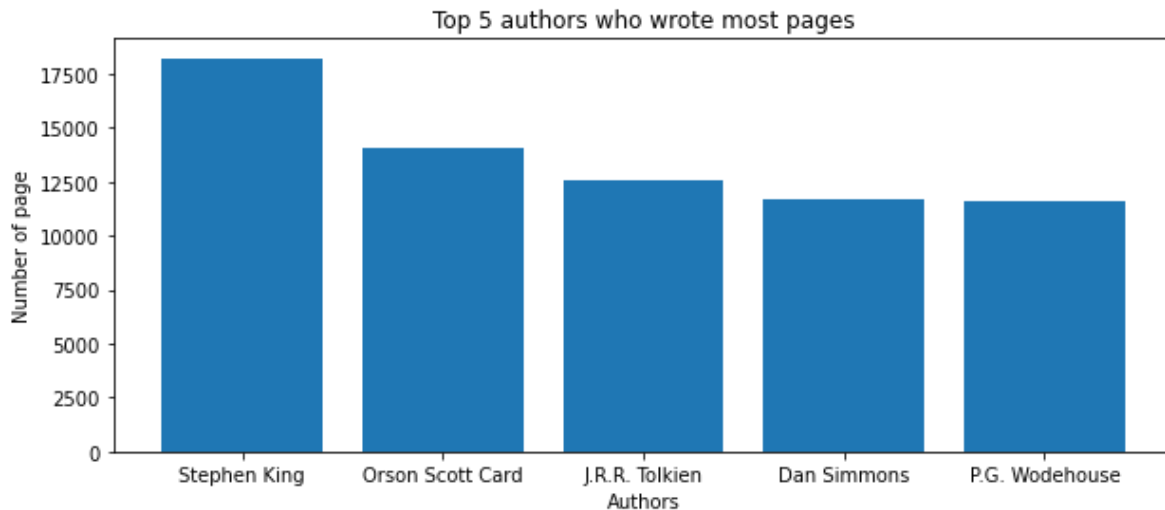
Out[26]:

	authors	num_page
5786	Stephen King	18219
4711	Orson Scott Card	14066
2640	J.R.R. Tolkien	12537
1120	Dan Simmons	11700
4747	P.G. Wodehouse	11619
...	...	...
1480	Douglas Preston/Lincoln Child/Scott Brick	0
4789	Patricia Cornwell/Kate Burton	0
2233	Gustave Flaubert/Donada Peters	0
2808	James Redfield/Lou Diamond Phillips	0
1238	David Bromberg/Jay Ungar/Molly Mason/Garrison ...	0

6639 rows × 2 columns

In [27]:

```
plt.figure(figsize=(10,4))  
plt.bar(df_sort["authors"][:5],height=df_sort["num_page"][:5])  
plt.xlabel("Authors")  
plt.ylabel("Number of page")  
plt.title("Top 5 authors who wrote most pages")  
plt.show()
```



In [ ]:

In [28]:

```
df2 =books_df.groupby("authors",as_index=False)["title"].count()
```

In [ ]:

In [29]:

```
df2 =pd.DataFrame(df2)
```

In [30]:

```
df2.sort_values(by=['title'],ascending=False)
```

Out[30]:

	authors	title
5786	Stephen King	40
4747	P.G. Wodehouse	40
5470	Rumiko Takahashi	39
4711	Orson Scott Card	35
46	Agatha Christie	33
...	...	...
3066	Joanne Harris/Fran Warde	1
811	Carolyn Keene/Mildred Benson	1
812	Carolyn Keene/Russell H. Tandy/Mildred Benson	1
3063	Joanna Weaver	1
6638	Éric-Emmanuel Schmitt	1

6639 rows × 2 columns

In [31]:

```
df3=pd.concat([df2,df["num_page"]],axis=1)
df3
```

Out[31]:

	authors	title	num_page
0	A.B. Yehoshua/Hillel Halkin	2	832
1	A.D.P. Briggs/Leo Tolstoy/Fyodor Dostoyevsky	1	256
2	A.E. Cunningham/Harlan Ellison/Charles F. Mill...	1	232
3	A.J. Jacobs	1	389
4	A.M. Homes	1	256
...	...	...	...
6634	Émile Zola/Henri Mitterand	1	462
6635	Émile Zola/Robert Lethbridge/Elinor Dorday	1	592
6636	Émile Zola/Robin Buss/Brian Nelson	1	438
6637	Émile Zola/Roger Pearson	2	1056
6638	Éric-Emmanuel Schmitt	1	85

6639 rows × 3 columns

In [32]:

```
df3["authors_avg_page"]=(df["num_page"]/df3["title"]).astype('int')
df3
```

Out[32]:

	authors	title	num_page	authors_avg_page
0	A.B. Yehoshua/Hillel Halkin	2	832	416
1	A.D.P. Briggs/Leo Tolstoy/Fyodor Dostoyevsky	1	256	256
2	A.E. Cunningham/Harlan Ellison/Charles F. Mill...	1	232	232
3	A.J. Jacobs	1	389	389
4	A.M. Homes	1	256	256
...	...	...	...	...
6634	Émile Zola/Henri Mitterand	1	462	462
6635	Émile Zola/Robert Lethbridge/Elinor Dorday	1	592	592
6636	Émile Zola/Robin Buss/Brian Nelson	1	438	438
6637	Émile Zola/Roger Pearson	2	1056	528
6638	Éric-Emmanuel Schmitt	1	85	85

6639 rows × 4 columns

In [33]:

```
df3.sort_values(by=["authors_avg_page"],ascending =False)
```

Out[33]:

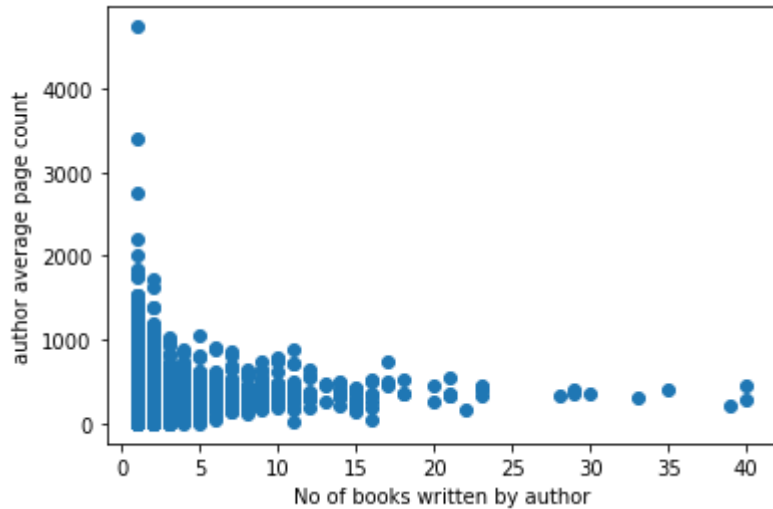
	authors	title	num_page	authors_avg_page
6564	Winston S. Churchill/John Keegan	1	4736	4736
3974	Marcel Proust/C.K. Scott Moncrieff/Frederick A...	1	3400	3400
1373	Dennis L. Kasper/Dan L. Longo/Stephen L. Hause...	1	2751	2751
336	Anonymous	1	2198	2198
3706	Laurence L. Brunton/John S. Lazo/Keith L. Parker	1	2021	2021
...	...	...	...	...
544	Barry Chubin	1	0	0
4133	Mark Twain/Michael Prichard	1	0	0
3162	John Grisham/Michael Beck/Frank Muller	1	0	0
3118	John Banville/John Lee	1	0	0
3161	John Grisham/Michael Beck/Blair Brown/D.W. Mof...	1	0	0

6639 rows × 4 columns

In [ ]:

In [34]:

```
plt.scatter(df3["title"],df3["authors_avg_page"])
plt.xlabel("No of books written by author")
plt.ylabel("author average page count")
plt.show()
```



In [35]:

```
# books written less than 200 pages
books_df[books_df["num_page"]<=200].info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2949 entries, 5 to 11117
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   bookID                 2949 non-null   int64
1   title                  2949 non-null   object
2   authors                2949 non-null   object
3   average_rating         2949 non-null   float64
4   isbn                   2949 non-null   object
5   isbn13                 2949 non-null   int64
6   language_code          2949 non-null   object
7   num_page               2949 non-null   int64
8   ratings_count          2949 non-null   int64
9   text_reviews_count     2949 non-null   int64
10  publication_date        2949 non-null   object
11  publisher               2949 non-null   object
12  w_rating                2949 non-null   float64
dtypes: float64(2), int64(5), object(6)
memory usage: 322.5+ KB
```

In [36]:

```
# Houghton Mifflin Harcourt most popular books
HMH_books = books_df[books_df["publisher"]=="Houghton Mifflin Harcourt"].sort_values(by=["w
HMH_books.head(3)
```

Out[36]:

	bookID	title	authors	average_rating	isbn	isbn13	language
	23	34	The Fellowship of the Ring (The Lord of the Ri...	J.R.R. Tolkien	4.36	0618346252	9780618346257
	4172	14942	Mrs. Dalloway	Virginia Woolf/Maureen Howard	3.79	0151009988	9780151009985
	1585	5472	Animal Farm / 1984	George Orwell/Christopher Hitchens	4.28	0151010269	9780151010264

In [37]:

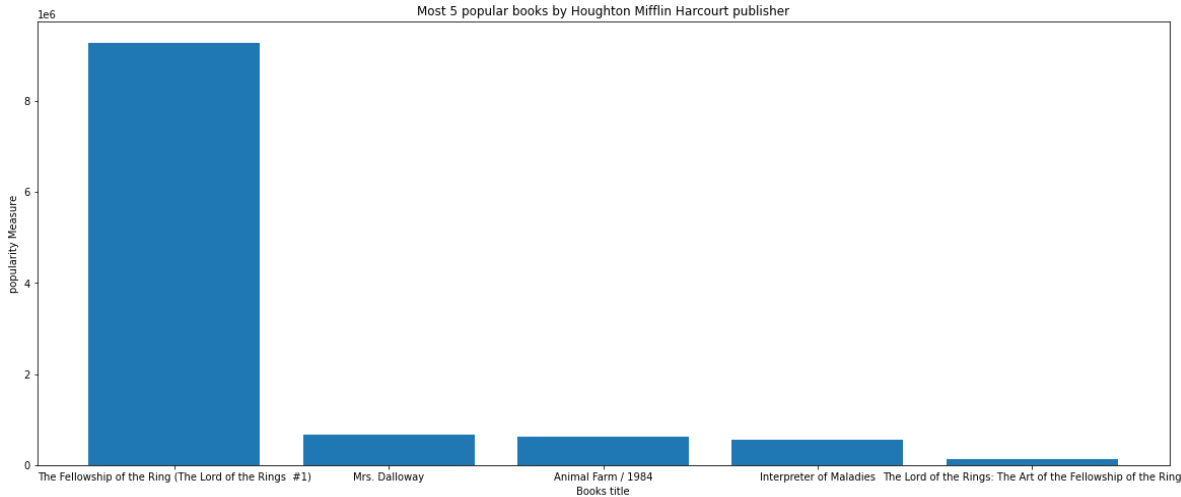
```
HMH_books.iloc[0].title
```

Out[37]:

'The Fellowship of the Ring (The Lord of the Rings #1)'

In [38]:

```
# 5 most popular books by HMH publishers
plt.figure(figsize=(20,8))
plt.bar(HMH_books["title"][:5],height=HMH_books["w_rating"][:5])
plt.xlabel("Books title")
plt.ylabel("popularity Measure")
plt.title('Most 5 popular books by Houghton Mifflin Harcourt publisher')
plt.show()
```





In [39]:

```
df5 =pd.DataFrame(books_df.groupby(by=["authors"],as_index=False)['w_rating'].max())
df5.head(3)
```

Out[39]:

	authors	w_rating
0	A.B. Yehoshua/Hillel Halkin	2918.93
1	A.D.P. Briggs/Leo Tolstoy/Fyodor Dostoyevsky	63.92
2	A.E. Cunningham/Harlan Ellison/Charles F. Mill...	53.95

In [40]:

```
df5[df5["authors"]=="Stephen King"]
```

Out[40]:

	authors	w_rating
5786	Stephen King	4129417.7

In [41]:

```
df5.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6639 entries, 0 to 6638
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   authors    6639 non-null   object
 1   w_rating    6639 non-null   float64
dtypes: float64(1), object(1)
memory usage: 103.9+ KB
```

In [42]:

```
# Least popular books of 90s
df6 =books_df[books_df["publication_date"]=="1990"].sort_values(by=["w_rating"])
df6.head(3)
```

Out[42]:

	bookID	title	authors	average_rating	isbn	isbn13	language_code
7287	28052	Dark Intimacy: Hope for Those in Difficult Pra...	David J. Hassel	4.00	0829407081	9780829407082	eng
10212	41404	The Savage Wars Of Peace: Soldiers' Voices 194...	Charles Allen	4.50	0718128826	9780718128821	eng
6310	23799	The Dark Descent Vol 1: The Color of Evil	David G. Hartwell	4.12	0246136677	9780246136671	eng



In [43]:

```
df6.iloc[0].title
```

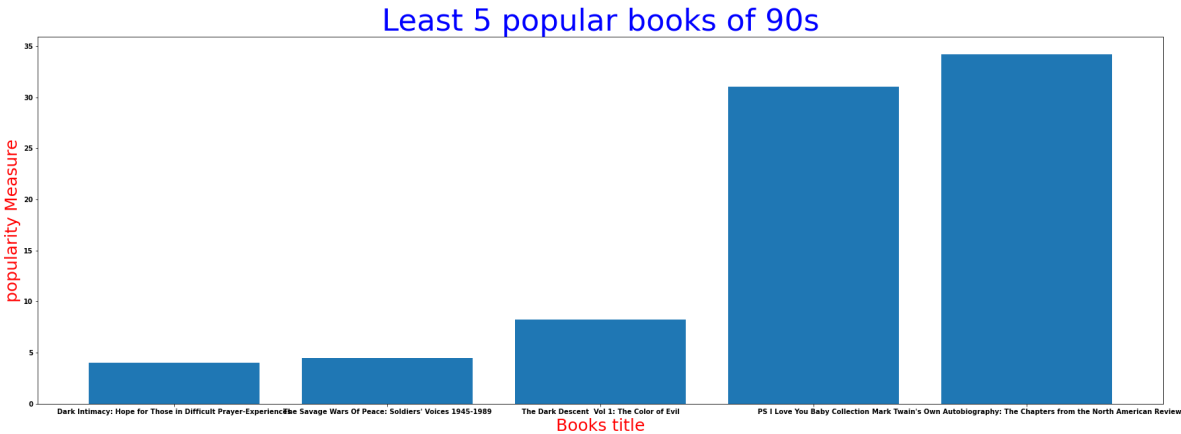
Out[43]:

'Dark Intimacy: Hope for Those in Difficult Prayer-Experiences'

In [44]:

```
plt.figure(figsize=(30,10))
plt.rcParams['font.size']=10
plt.rcParams['font.weight']='bold'
plt.bar(df6["title"][:5],height=df6["w_rating"][:5])
plt.xlabel("Books title",fontsize=25, color='r')
plt.ylabel("popularity Measure",fontsize=25, color='r')
plt.title('Least 5 popular books of 90s',fontsize=45,color='b')

plt.show()
```



In [45]:

```
# books rated highest which have pages more than 500
df7 =books_df[books_df["num_page"]>=500].sort_values(by=["average_rating"],ascending=False)
```

In [46]:

```
df7.head(2)
```

Out[46]:

bookID		title	authors	average_rating	isbn	isbn13	language_code
5648	20957	Colossians and Philemon: A Critical and Exeget...	R. McL. Wilson	5.00	0567044718	9780567044716	eng
6587	24812	The Complete Calvin and Hobbes	Bill Watterson	4.82	0740748475	9780740748479	eng

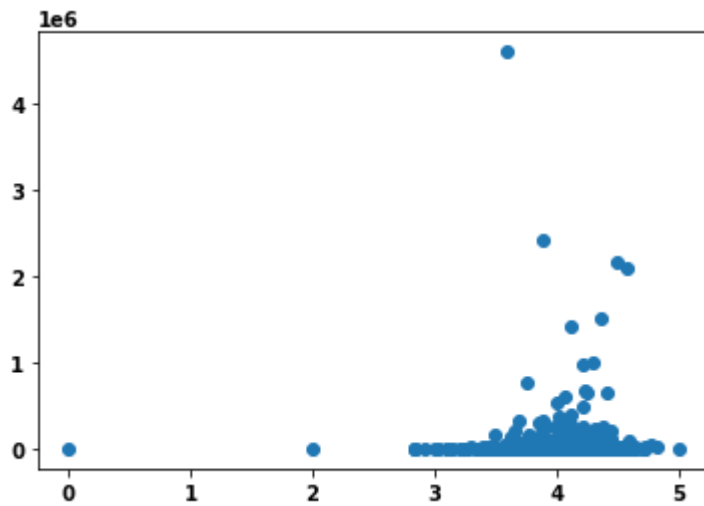


In [47]:

```
plt.scatter(df7["average_rating"],df7["ratings_count"])
```

Out[47]:

<matplotlib.collections.PathCollection at 0x17d7ae5b160>



In [48]:

```
df7.iloc[0].title
```

Out[48]:

'Colossians and Philemon: A Critical and Exegetical Commentary (International Critical Commentary)'

In [49]:

```
corr = books_df.corr()  
corr
```

Out[49]:

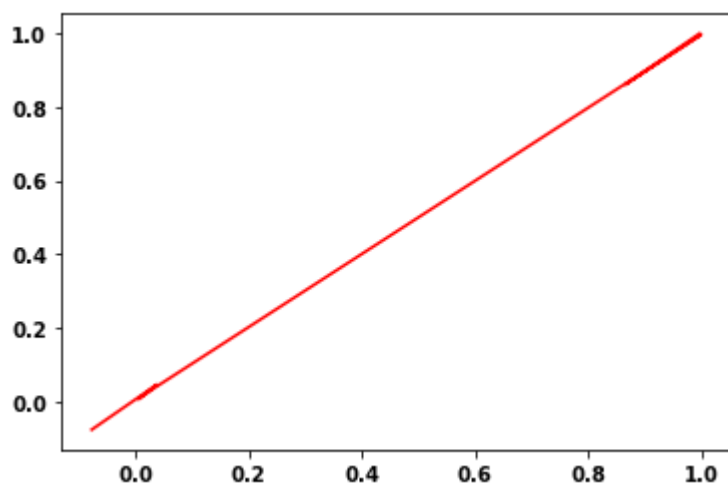
	bookID	average_rating	isbn13	num_page	ratings_count	text_reviews
bookID	1.000000	-0.036207	-0.002185	-0.026134	-0.076820	-0.
average_rating	-0.036207	1.000000	-0.001965	0.150477	0.038224	0.
isbn13	-0.002185	-0.001965	1.000000	-0.009818	0.005494	0.
num_page	-0.026134	0.150477	-0.009818	1.000000	0.034337	0.
ratings_count	-0.076820	0.038224	0.005494	0.034337	1.000000	0.
text_reviews_count	-0.081289	0.033663	0.008152	0.036977	0.865977	1.
w_rating	-0.077622	0.044985	0.005294	0.036182	0.997043	0.

In [50]:

```
plt.plot(corr["ratings_count"],corr["w_rating"],color='r')
```

Out[50]:

```
[<matplotlib.lines.Line2D at 0x17d7aecc640>]
```

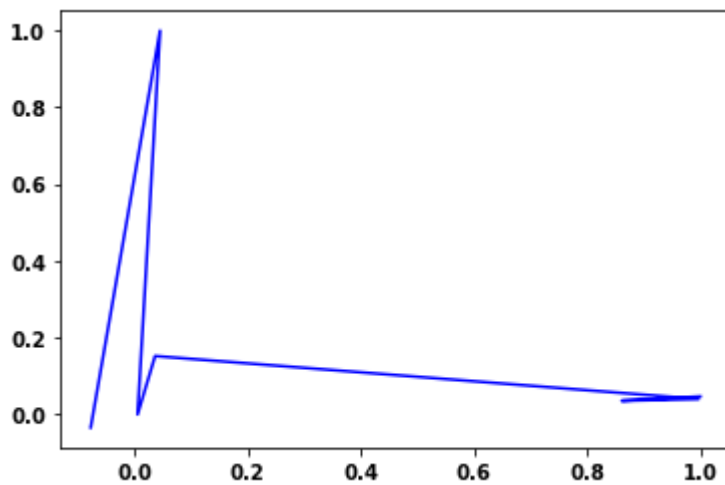


In [51]:

```
plt.plot(corr["w_rating"],corr["average_rating"],color='b')
```

Out[51]:

[<matplotlib.lines.Line2D at 0x17d7af27cd0>]

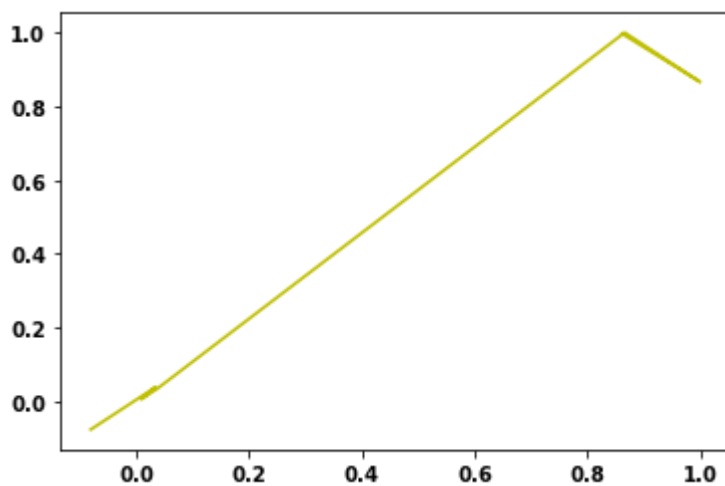


In [52]:

```
plt.plot(corr["text_reviews_count"],corr["ratings_count"],color='y')
```

Out[52]:

[<matplotlib.lines.Line2D at 0x17d7af8abb0>]



In [53]:

```
np.std(books_df,axis=0)
```

```
C:\Users\RAHUL SUTHAR\anaconda3\lib\site-packages\numpy\core\fromnumeric.py:3571: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
```

```
    return std(axis=axis, dtype=dtype, out=out, ddof=ddof, **kwargs)
```

Out[53]:

bookID	1.309414e+04
average_rating	3.504696e-01
isbn13	4.429559e+11
num_page	2.411418e+02
ratings_count	1.124941e+05
text_reviews_count	2.576504e+03
w_rating	4.532409e+05
dtype:	float64

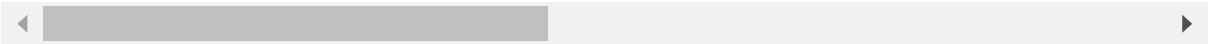
In [54]:

```
books_df.sort_values(by=['average_rating'],ascending=False)
```

Out[54]:

bookID		title	authors	average_rating	isbn
624	2034	Comoediae 1: Acharenses/Equites/Nubes/Vespae/P...	Aristophanes/F.W. Hall/W.M. Geldart	5.0	0198145047
9893	39829	His Princess Devotional: A Royal Encounter Wit...	Sheri Rose Shepherd	5.0	1590529626
4788	17224	The Diamond Color Meditation: Color Pathway to...	John Diamond	5.0	1890995525
9324	36853	Tyrannosaurus Wrecks (Stanley #1)	Laura Driscoll/Alisa Klayman-Grodsky/Eric ...	5.0	0786845031
9720	38804	The Irish Anatomist: A Study of Flann O'Brien	Keith Donohue	5.0	1930901356
...	...	...	...	...	...
6877	25903	I'll Be Home Before Midnight and I Won't Get P...	Anthony E. Wolf	0.0	0394755669
3492	12712	Brodie's notes on Aldous Huxley's brave new world	Graham Handley	0.0	0333581296
2533	9338	Insights: Talks On The Nature Of Existence	Frederick P. Lenz	0.0	1932206086
2532	9337	How To Meditate: An Anthology Of Talks On Medi...	Frederick P. Lenz	0.0	1932206108
10497	42758	The Fantastic Vampire: Studies in the Children...	James Craig Holte	0.0	0313309337

11123 rows × 13 columns





In [55]:

```
books_df.iloc[5562]
```

Out[55]:

```
bookID                20315
title                The Dark City (Eliot Ness #1)
authors              Max Allan Collins
average_rating        3.88
isbn                 0727862421
isbn13              9780727862426
language_code        eng
num_page             233
ratings_count        49
text_reviews_count    5
publication_date      2005
publisher            Severn House Publishers
w_rating             190.12
Name: 5562, dtype: object
```

In [56]:

```
11123/2
```

Out[56]:

```
5561.5
```

In [57]:

```
books_df[books_df["authors"]=="Stephen King"].count()
```

Out[57]:

```
bookID                40
title                40
authors              40
average_rating        40
isbn                 40
isbn13              40
language_code        40
num_page             40
ratings_count        40
text_reviews_count    40
publication_date      40
publisher            40
w_rating             40
dtype: int64
```

In [ ]:

In [ ]:

We have given a books dataset. Dataset contains the following columns

- ['bookID', 'title', 'authors', 'average\_rating', 'isbn', 'isbn13', 'language\_code', 'num\_pages', 'ratings\_count', 'text\_reviews\_count', 'publication\_date', 'publisher']
- Data type of columns is int, object or float type. Dataset not contains any null value but there are one bad lines which we have to skip.
- Shape of Dataset is (11123, 12) means there are total 11123 rows and 12 columns. There are also some outliers like some books with 0 average rating because rating count is also 0. also 1 row which has 5 rating but rating count is 0. and some rows with 5 rating with -12 rating counts which are outliers.
- The size of dataset is 133476.
- To measure popularity of books we can't use only average rating or only rating count so we use multiple of average rating and rating count. using this we got Most Popular books as "Twilight (Twilight #1)" by "Stephenie Meyer" with average rating of 3.59
- Book published by Little Brown and Company in 9 June 2006.
- Books which are more pages rated mostly higher.
- In 60s only 6 books were published and most popular books of 60s is "Paws Before Dying (A Dog Lover's Mystery #4)" written by "Susan Conant" and published by "Berkley Books".
- The author "Stephen King" wrote most pages. 18219 pages are written by Stephen King. He writes 40 books.
- 2949 books that are written in less or equal 200 pages.
- Most popular books which are published by Houghton Mifflin Harcourt is "The Fellowship of the Ring (The Lord of the Rings #1)" written by J.R.R. Tolkien and published in 2003.
- "Dark Intimacy: Hope for Those in Difficult Prayer-Experiences" is least popular book of 90s.
- "Colossians and Philemon: A Critical and Exegetical Commentary (International Critical Commentary)" is highest rated books which have more than 500 pages written by "R. McL. Wilson" but consider the fact that rating count is only 1.

In [ ]: