

April 25, 2024

```
[2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
[4]: df = pd.read_csv("books.csv")
df
```

```
[4]:
```

	book_id	goodreads_book_id	best_book_id	work_id	books_count	\
0	1	2767052	2767052	2792775	272	
1	2	3	3	4640799	491	
2	3	41865	41865	3212258	226	
3	6	11870085	11870085	16827462	226	
4	12	13335037	13335037	13155899	210	
...	
1349	9925	86737	86737	3877968	52	
1350	9937	13010211	13010211	18171867	22	
1351	9942	16074758	16074758	21869436	18	
1352	9947	21393526	21393526	40690062	19	
1353	9955	13065327	13065327	18230950	25	

	isbn	isbn13	authors	\
0	439023483	9.780439e+12	Suzanne Collins	
1	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	
2	316015849	9.780316e+12	Stephenie Meyer	
3	525478817	9.780525e+12	John Green	
4	62024035	9.780062e+12	Veronica Roth	
...	
1349	1582349177	9.781582e+12	Mary Hoffman	
1350	1596435712	9.781596e+12	Caragh M. O'Brien	
1351	1442486597	9.781442e+12	Abigail Haas, Abby McDonald	
1352	62320521	9.780062e+12	Maria Dahvana Headley	
1353	802734375	9.780803e+12	Simone Elkeles	

	original_publication_year	original_title	\
0	2008.0	The Hunger Games	
1	1997.0	Harry Potter and the Philosopher's Stone	
2	2005.0	Twilight	

3	2012.0	The Fault in Our Stars
4	2011.0	Divergent
...
1349	2002.0	City of Masks
1350	2012.0	Promised
1351	2013.0	Dangerous Girls
1352	2015.0	Magonia
1353	2013.0	Wild Cards

	...	ratings_count	work_ratings_count	work_text_reviews_count	\
0	...	4780653	4942365	155254	
1	...	4602479	4800065	75867	
2	...	3866839	3916824	95009	
3	...	2346404	2478609	140739	
4	...	1903563	2216814	101023	
...	
1349	...	12048	13385	555	
1350	...	11766	12884	1260	
1351	...	10439	12970	2631	
1352	...	12510	13652	2910	
1353	...	13954	15400	1514	

	ratings_1	ratings_2	ratings_3	ratings_4	ratings_5	\
0	66715	127936	560092	1481305	2706317	
1	75504	101676	455024	1156318	3011543	
2	456191	436802	793319	875073	1355439	
3	47994	92723	327550	698471	1311871	
4	36315	82870	310297	673028	1114304	
...	
1349	314	758	3154	4872	4287	
1350	256	1098	3565	4392	3573	
1351	203	553	2029	4567	5618	
1352	577	1440	3881	4772	2982	
1353	406	965	3331	5038	5660	

	image_url	\
0	https://images.gr-assets.com/books/1447303603m...	
1	https://images.gr-assets.com/books/1474154022m...	
2	https://images.gr-assets.com/books/1361039443m...	
3	https://images.gr-assets.com/books/1360206420m...	
4	https://images.gr-assets.com/books/1328559506m...	
...	...	
1349	https://images.gr-assets.com/books/1294870497m...	
1350	https://images.gr-assets.com/books/1330546217m...	
1351	https://images.gr-assets.com/books/1356513050m...	
1352	https://images.gr-assets.com/books/1413479866m...	
1353	https://images.gr-assets.com/books/1420836321m...	

```

                                small_image_url
0    https://images.gr-assets.com/books/1447303603s...
1    https://images.gr-assets.com/books/1474154022s...
2    https://images.gr-assets.com/books/1361039443s...
3    https://images.gr-assets.com/books/1360206420s...
4    https://images.gr-assets.com/books/1328559506s...
...
1349 https://images.gr-assets.com/books/1294870497s...
1350 https://images.gr-assets.com/books/1330546217s...
1351 https://images.gr-assets.com/books/1356513050s...
1352 https://images.gr-assets.com/books/1413479866s...
1353 https://images.gr-assets.com/books/1420836321s...

```

[1354 rows x 23 columns]

```

[5]: missing_values = df.isna()
      print(missing_values)
      Cleandf = df.dropna()

```

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	\
0	False	False	False	False	False	False	
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
3	False	False	False	False	False	False	
4	False	False	False	False	False	False	
...	
1349	False	False	False	False	False	False	
1350	False	False	False	False	False	False	
1351	False	False	False	False	False	False	
1352	False	False	False	False	False	False	
1353	False	False	False	False	False	False	

	isbn13	authors	original_publication_year	original_title	...	\
0	False	False	False	False	...	
1	False	False	False	False	...	
2	False	False	False	False	...	
3	False	False	False	False	...	
4	False	False	False	False	...	
...	
1349	False	False	False	False	...	
1350	False	False	False	False	...	
1351	False	False	False	False	...	
1352	False	False	False	False	...	
1353	False	False	False	False	...	

	ratings_count	work_ratings_count	work_text_reviews_count	ratings_1	\
--	---------------	--------------------	-------------------------	-----------	---

0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...
1349	False	False	False	False
1350	False	False	False	False
1351	False	False	False	False
1352	False	False	False	False
1353	False	False	False	False

	ratings_2	ratings_3	ratings_4	ratings_5	image_url	small_image_url
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...
1349	False	False	False	False	False	False
1350	False	False	False	False	False	False
1351	False	False	False	False	False	False
1352	False	False	False	False	False	False
1353	False	False	False	False	False	False

[1354 rows x 23 columns]

```
[6]: harry_potter_books = Cleandf[Cleandf["title"].str.contains("Harry Potter",
↪case=False, na=False)]
print(harry_potter_books)
```

	book_id	goodreads_book_id	best_book_id	work_id	books_count	\
1	2	3	3	4640799	491	
6	18	5	5	2402163	376	
8	21	2	2	2809203	307	
9	23	15881	15881	6231171	398	
10	24	6	6	3046572	332	
11	25	136251	136251	2963218	263	
12	27	1	1	41335427	275	
96	422	862041	862041	2962492	76	
613	3753	10	10	21457570	6	
1036	7018	483445	483445	471792	42	

	isbn	isbn13	authors	\
1	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	
6	043965548X	9.780440e+12	J.K. Rowling, Mary GrandPré, Rufus Beck	
8	439358078	9.780439e+12	J.K. Rowling, Mary GrandPré	
9	439064864	9.780439e+12	J.K. Rowling, Mary GrandPré	

10	439139600	9.780439e+12	J.K. Rowling, Mary GrandPré
11	545010225	9.780545e+12	J.K. Rowling, Mary GrandPré
12	439785960	9.780440e+12	J.K. Rowling, Mary GrandPré
96	545044251	9.780545e+12	J.K. Rowling
613	439827604	9.780440e+12	J.K. Rowling
1036	042519891X	9.780425e+12	David Colbert

	original_publication_year \
1	1997.0
6	1999.0
8	2003.0
9	1998.0
10	2000.0
11	2007.0
12	2005.0
96	1998.0
613	2005.0
1036	2001.0

	original_title	... ratings_count \
1	Harry Potter and the Philosopher's Stone	... 4602479
6	Harry Potter and the Prisoner of Azkaban	... 1832823
8	Harry Potter and the Order of the Phoenix	... 1735368
9	Harry Potter and the Chamber of Secrets	... 1779331
10	Harry Potter and the Goblet of Fire	... 1753043
11	Harry Potter and the Deathly Hallows	... 1746574
12	Harry Potter and the Half-Blood Prince	... 1678823
96	Complete Harry Potter Boxed Set	... 190050
613	Harry Potter Collection (Harry Potter, #1-6)	... 24618
1036	The Magical Worlds of Harry Potter: A Treasury...	... 13820

	work_ratings_count	work_text_reviews_count	ratings_1	ratings_2 \
1	4800065	75867	75504	101676
6	1969375	36099	6716	20413
8	1840548	28685	9528	31577
9	1906199	34172	8253	42251
10	1868642	31084	6676	20210
11	1847395	51942	9363	22245
12	1785676	27520	7308	21516
96	204125	6508	1105	1285
613	26274	882	203	186
1036	15145	267	329	1125

	ratings_3	ratings_4	ratings_5 \
1	455024	1156318	3011543
6	166129	509447	1266670
8	180210	494427	1124806
9	242345	548266	1065084

10	151785	494926	1195045
11	113646	383914	1318227
12	136333	459028	1161491
96	7020	30666	164049
613	946	3891	21048
1036	3766	3593	6332

	image_url \
1	https://images.gr-assets.com/books/1474154022m...
6	https://images.gr-assets.com/books/1499277281m...
8	https://images.gr-assets.com/books/1387141547m...
9	https://images.gr-assets.com/books/1474169725m...
10	https://images.gr-assets.com/books/1361482611m...
11	https://images.gr-assets.com/books/1474171184m...
12	https://images.gr-assets.com/books/1361039191m...
96	https://images.gr-assets.com/books/1392579059m...
613	https://images.gr-assets.com/books/1328867351m...
1036	https://s.gr-assets.com/assets/nophoto/book/11...

	small_image_url
1	https://images.gr-assets.com/books/1474154022s...
6	https://images.gr-assets.com/books/1499277281s...
8	https://images.gr-assets.com/books/1387141547s...
9	https://images.gr-assets.com/books/1474169725s...
10	https://images.gr-assets.com/books/1361482611s...
11	https://images.gr-assets.com/books/1474171184s...
12	https://images.gr-assets.com/books/1361039191s...
96	https://images.gr-assets.com/books/1392579059s...
613	https://images.gr-assets.com/books/1328867351s...
1036	https://s.gr-assets.com/assets/nophoto/book/50...

[10 rows x 23 columns]

```
[7]: shp = harry_potter_books[
      [
        "book_id",
        "books_count",
        "original_title",
        "average_rating",
        "ratings_count",
      ]
    ]
shp.head(10)
```

[7]:	book_id	books_count	original_title \
1	2	491	Harry Potter and the Philosopher's Stone
6	18	376	Harry Potter and the Prisoner of Azkaban

8	21	307	Harry Potter and the Order of the Phoenix
9	23	398	Harry Potter and the Chamber of Secrets
10	24	332	Harry Potter and the Goblet of Fire
11	25	263	Harry Potter and the Deathly Hallows
12	27	275	Harry Potter and the Half-Blood Prince
96	422	76	Complete Harry Potter Boxed Set
613	3753	6	Harry Potter Collection (Harry Potter, #1-6)
1036	7018	42	The Magical Worlds of Harry Potter: A Treasury...

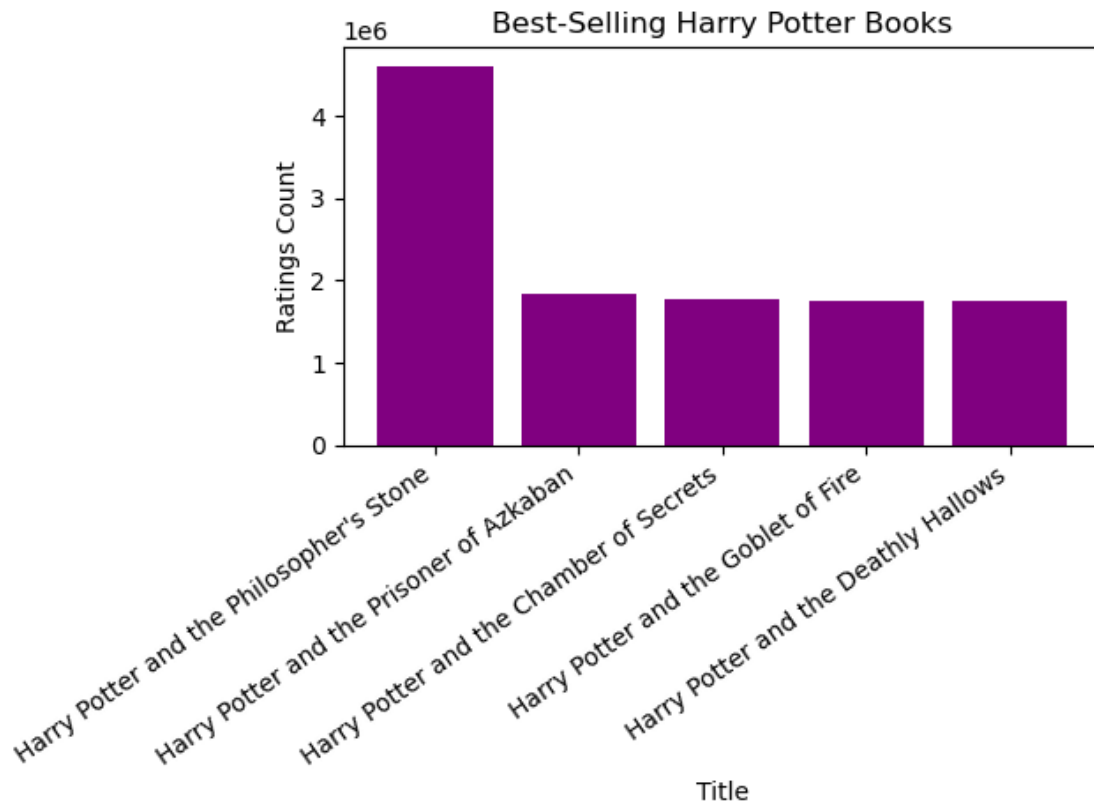
	average_rating	ratings_count
1	4.44	4602479
6	4.53	1832823
8	4.46	1735368
9	4.37	1779331
10	4.53	1753043
11	4.61	1746574
12	4.54	1678823
96	4.74	190050
613	4.73	24618
1036	3.96	13820

```
[8]: harry_potter_books_sorted = shp.sort_values(by="ratings_count", ascending=False)
best_selling_harry_potter_books = harry_potter_books_sorted.head(5)
print("Top {} best-selling Harry Potter books:".format(5))
print(best_selling_harry_potter_books[["original_title", "ratings_count"]])
```

Top 5 best-selling Harry Potter books:

	original_title	ratings_count
1	Harry Potter and the Philosopher's Stone	4602479
6	Harry Potter and the Prisoner of Azkaban	1832823
9	Harry Potter and the Chamber of Secrets	1779331
10	Harry Potter and the Goblet of Fire	1753043
11	Harry Potter and the Deathly Hallows	1746574

```
[9]: plt.bar(
best_selling_harry_potter_books["original_title"],
best_selling_harry_potter_books["ratings_count"],
color="purple",
) # Create bar plot
plt.xlabel("Title")
plt.ylabel("Ratings Count")
plt.title("Best-Selling Harry Potter Books")
plt.xticks(rotation=35, ha="right") # Rotate x-axis labels for better_visibility
plt.tight_layout() # Adjust spacing
plt.show()
```



```
[10]: average_rating_harry_potter = harry_potter_books["average_rating"].mean()  
print("\nAverage rating for Harry Potter books: {:.2f}".  
      ↪format(average_rating_harry_potter))
```

Average rating for Harry Potter books: 4.49