

# untitled

April 24, 2024

```
[1]: import pandas as pd
import datetime as dt
import numpy as np
import warnings
```

```
[2]: df = pd.read_csv('books.csv')
df
```

```
[2]:
```

	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	<a href="https://images.gr-assets.com/books/1447303603m...">https://images.gr-assets.com/books/1447303603m...</a>	
1	<a href="https://images.gr-assets.com/books/1474154022m...">https://images.gr-assets.com/books/1474154022m...</a>	
2	<a href="https://images.gr-assets.com/books/1361039443m...">https://images.gr-assets.com/books/1361039443m...</a>	
3	<a href="https://images.gr-assets.com/books/1360206420m...">https://images.gr-assets.com/books/1360206420m...</a>	
4	<a href="https://images.gr-assets.com/books/1328559506m...">https://images.gr-assets.com/books/1328559506m...</a>	
...	...	
1349	<a href="https://images.gr-assets.com/books/1294870497m...">https://images.gr-assets.com/books/1294870497m...</a>	
1350	<a href="https://images.gr-assets.com/books/1330546217m...">https://images.gr-assets.com/books/1330546217m...</a>	
1351	<a href="https://images.gr-assets.com/books/1356513050m...">https://images.gr-assets.com/books/1356513050m...</a>	
1352	<a href="https://images.gr-assets.com/books/1413479866m...">https://images.gr-assets.com/books/1413479866m...</a>	

```

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]
```

```
[4]: df.tail()
      df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1354 entries, 0 to 1353
Data columns (total 23 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   book_id                               1354 non-null   int64
1   goodreads_book_id                    1354 non-null   int64
2   best_book_id                         1354 non-null   int64
3   work_id                              1354 non-null   int64
4   books_count                          1354 non-null   int64
5   isbn                                  1302 non-null   object
6   isbn13                               1310 non-null   float64
7   authors                              1354 non-null   object
8   original_publication_year            1351 non-null   float64
9   original_title                       1302 non-null   object
10  title                                1354 non-null   object
11  language_code                        1245 non-null   object
12  average_rating                       1354 non-null   float64
13  ratings_count                        1354 non-null   int64
14  work_ratings_count                   1354 non-null   int64
15  work_text_reviews_count              1354 non-null   int64
16  ratings_1                            1354 non-null   int64
17  ratings_2                            1354 non-null   int64
18  ratings_3                            1354 non-null   int64
19  ratings_4                            1354 non-null   int64
20  ratings_5                            1354 non-null   int64
21  image_url                            1354 non-null   object

```

```

22 small_image_url          1354 non-null    object
dtypes: float64(3), int64(13), object(7)
memory usage: 243.4+ KB

```

```

[5]: df['original_publication_year'] = pd.
      ↪to_datetime(df['original_publication_year'])
      print(df['original_publication_year'])

```

```

0      1970-01-01 00:00:00.0000002008
1      1970-01-01 00:00:00.0000001997
2      1970-01-01 00:00:00.0000002005
3      1970-01-01 00:00:00.0000002012
4      1970-01-01 00:00:00.0000002011
...
1349   1970-01-01 00:00:00.0000002002
1350   1970-01-01 00:00:00.0000002012
1351   1970-01-01 00:00:00.0000002013
1352   1970-01-01 00:00:00.0000002015
1353   1970-01-01 00:00:00.0000002013
Name: original_publication_year, Length: 1354, dtype: datetime64[ns]

```

```

[6]: df = df[(df.books_count>=0)&(df.ratings_count>=0)&(df.
      ↪work_ratings_count>=0)&(df.work_text_reviews_count>=0)&(df.ratings_1>=0)&(df.
      ↪ratings_2>=0)&(df.ratings_3>=0)&(df.ratings_4>=0)&(df.ratings_5>=0)]
      print(df)

```

	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	1970-01-01 00:00:00.000002008	The Hunger Games
1	1970-01-01 00:00:00.000001997	Harry Potter and the Philosopher's Stone
2	1970-01-01 00:00:00.000002005	Twilight
3	1970-01-01 00:00:00.000002012	The Fault in Our Stars
4	1970-01-01 00:00:00.000002011	Divergent
...	...	...
1349	1970-01-01 00:00:00.000002002	City of Masks
1350	1970-01-01 00:00:00.000002012	Promised
1351	1970-01-01 00:00:00.000002013	Dangerous Girls
1352	1970-01-01 00:00:00.000002015	Magonia
1353	1970-01-01 00:00:00.000002013	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	<a href="https://images.gr-assets.com/books/1447303603m...">https://images.gr-assets.com/books/1447303603m...</a>
1	<a href="https://images.gr-assets.com/books/1474154022m...">https://images.gr-assets.com/books/1474154022m...</a>
2	<a href="https://images.gr-assets.com/books/1361039443m...">https://images.gr-assets.com/books/1361039443m...</a>
3	<a href="https://images.gr-assets.com/books/1360206420m...">https://images.gr-assets.com/books/1360206420m...</a>

```

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]

```
[7]: df = df.drop_duplicates()
      print(df)
```

	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	1970-01-01 00:00:00.000002008	The Hunger Games
1	1970-01-01 00:00:00.000001997	Harry Potter and the Philosopher's Stone
2	1970-01-01 00:00:00.000002005	Twilight
3	1970-01-01 00:00:00.000002012	The Fault in Our Stars
4	1970-01-01 00:00:00.000002011	Divergent
...	...	...
1349	1970-01-01 00:00:00.000002002	City of Masks
1350	1970-01-01 00:00:00.000002012	Promised
1351	1970-01-01 00:00:00.000002013	Dangerous Girls
1352	1970-01-01 00:00:00.000002015	Magonia
1353	1970-01-01 00:00:00.000002013	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	<a href="https://images.gr-assets.com/books/1447303603m...">https://images.gr-assets.com/books/1447303603m...</a>
1	<a href="https://images.gr-assets.com/books/1474154022m...">https://images.gr-assets.com/books/1474154022m...</a>
2	<a href="https://images.gr-assets.com/books/1361039443m...">https://images.gr-assets.com/books/1361039443m...</a>
3	<a href="https://images.gr-assets.com/books/1360206420m...">https://images.gr-assets.com/books/1360206420m...</a>

```

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]

```
[8]: df.isnull().sum()
```

```

[8]: book_id                0
    goodreads_book_id       0
    best_book_id            0
    work_id                 0
    books_count             0
    isbn                   52
    isbn13                 44
    authors                0
    original_publication_year  3
    original_title         52
    title                  0
    language_code         109
    average_rating          0
    ratings_count          0
    work_ratings_count      0
    work_text_reviews_count  0
    ratings_1              0
    ratings_2              0
    ratings_3              0
    ratings_4              0
    ratings_5              0
    image_url              0

```



```
small_image_url          0
dtype: int64
```

```
[9]: df.dropna(inplace=True)
df.isnull().sum()
```

```
[9]: book_id          0
goodreads_book_id    0
best_book_id         0
work_id              0
books_count          0
isbn                 0
isbn13               0
authors              0
original_publication_year  0
original_title        0
title                0
language_code         0
average_rating        0
ratings_count         0
work_ratings_count    0
work_text_reviews_count  0
ratings_1             0
ratings_2             0
ratings_3             0
ratings_4             0
ratings_5             0
image_url             0
small_image_url       0
dtype: int64
```

```
[10]: harry_potter_df = df[df['original_title'].str.contains('Harry Potter',
↪case=False)]
print (harry_potter_df)
```

	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	1970-01-01 00:00:00.000001997
6	1970-01-01 00:00:00.000001999
8	1970-01-01 00:00:00.000002003
9	1970-01-01 00:00:00.000001998
10	1970-01-01 00:00:00.000002000
11	1970-01-01 00:00:00.000002007
12	1970-01-01 00:00:00.000002005
96	1970-01-01 00:00:00.000001998
613	1970-01-01 00:00:00.000002005
1036	1970-01-01 00:00:00.000002001

	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]

```
[11]: MostSellingBooks = harry_potter_df.loc[harry_potter_df['best_book_id'].idxmax()]
print (MostSellingBooks)
```

book_id	422
goodreads_book_id	862041
best_book_id	862041
work_id	2962492
books_count	76
isbn	545044251
isbn13	9780545044260.0

authors	J.K. Rowling
original_publication_year	1970-01-01 00:00:00.000001998
original_title	Complete Harry Potter Boxed Set
title	Harry Potter Boxset (Harry Potter, #1-7)
language_code	eng
average_rating	4.74
ratings_count	190050
work_ratings_count	204125
work_text_reviews_count	6508
ratings_1	1105
ratings_2	1285
ratings_3	7020
ratings_4	30666
ratings_5	164049
image_url	<a href="https://images.gr-assets.com/books/1392579059m...">https://images.gr-assets.com/books/1392579059m...</a>
small_image_url	<a href="https://images.gr-assets.com/books/1392579059s...">https://images.gr-assets.com/books/1392579059s...</a>

Name: 96, dtype: object

```
[12]: AverageRating = harry_potter_df['average_rating'].mean()
      print (AverageRating)
```

4.4910000000000005

```
[ ]:
```