

```
import pandas as pd
import numpy as np
df=pd.read_csv('books.csv')
```

```
df.head()
```

	book_id	goodreads_book_id	best_book_id	work_id	books_count
isbn \					
0	1	2767052	2767052	2792775	272
439023483					
1	2	3	3	4640799	491
439554934					
2	3	41865	41865	3212258	226
316015849					
3	6	11870085	11870085	16827462	226
525478817					
4	12	13335037	13335037	13155899	210
62024035					

	isbn13	authors
original_publication_year \		
0	9.780439e+12	Suzanne Collins
2008.0		
1	9.780440e+12	J.K. Rowling, Mary GrandPré
1997.0		
2	9.780316e+12	Stephenie Meyer
2005.0		
3	9.780525e+12	John Green
2012.0		
4	9.780062e+12	Veronica Roth
2011.0		

	original_title	... ratings_count \
0	The Hunger Games	... 4780653
1	Harry Potter and the Philosopher's Stone	... 4602479
2	Twilight	... 3866839
3	The Fault in Our Stars	... 2346404
4	Divergent	... 1903563

	work_ratings_count	work_text_reviews_count	ratings_1	ratings_2 \
0	4942365	155254	66715	127936
1	4800065	75867	75504	101676
2	3916824	95009	456191	436802
3	2478609	140739	47994	92723
4	2216814	101023	36315	82870

	ratings_3	ratings_4	ratings_5 \
0	560092	1481305	2706317
1	455024	1156318	3011543
2	793319	875073	1355439

3	327550	698471	1311871
4	310297	673028	1114304

	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...

	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...

[5 rows x 23 columns]

df.describe()

	book_id	goodreads_book_id	best_book_id	work_id \
count	1354.000000	1.354000e+03	1.354000e+03	1.354000e+03
mean	4453.584195	5.951852e+06	6.120589e+06	8.707028e+06
std	2894.277455	6.664595e+06	6.935008e+06	9.813696e+06
min	1.000000	1.000000e+00	1.000000e+00	1.150000e+02
25%	1860.250000	1.537868e+05	1.537962e+05	1.375035e+06
50%	4177.500000	3.305318e+06	3.422646e+06	4.005716e+06
75%	6814.500000	9.917380e+06	1.019388e+07	1.435717e+07
max	9955.000000	3.207567e+07	3.360215e+07	4.963819e+07

	books_count	isbn13	original_publication_year
average_rating \			
count	1354.000000	1.310000e+03	1351.000000
1354.000000			
mean	50.330871	9.766700e+12	2003.422650
3.999357			
std	61.338867	3.572069e+11	16.779301
0.224263			
min	1.000000	7.678361e+10	1868.000000
3.230000			
25%	22.000000	9.780152e+12	2003.000000
3.850000			
50%	37.000000	9.780440e+12	2008.000000
4.000000			
75%	58.000000	9.780805e+12	2011.000000
4.160000			
max	1314.000000	9.788424e+12	2017.000000
4.740000			

	ratings_count	work_ratings_count	work_text_reviews_count	\
count	1.354000e+03	1.354000e+03	1354.000000	
mean	9.160429e+04	9.915569e+04	5151.093058	
std	2.871266e+05	3.023637e+05	10730.335273	
min	6.221000e+03	8.833000e+03	49.000000	
25%	1.759325e+04	1.918150e+04	1162.500000	
50%	2.943000e+04	3.255150e+04	2208.000000	
75%	6.073800e+04	6.681275e+04	4690.750000	
max	4.780653e+06	4.942365e+06	155254.000000	

	ratings_1	ratings_2	ratings_3	ratings_4	ratings_5
count	1354.000000	1354.000000	1354.000000	1.354000e+03	1.354000e+03
mean	2297.409158	5005.615953	17528.918021	3.060591e+04	4.371784e+04
std	13708.507239	16259.838433	43549.306920	8.427851e+04	1.610638e+05
min	33.000000	133.000000	826.000000	1.660000e+03	2.005000e+03
25%	306.000000	978.000000	4140.500000	6.360500e+03	6.981500e+03
50%	619.000000	1732.500000	6557.000000	1.079550e+04	1.182650e+04
75%	1355.000000	3644.500000	13312.250000	2.227500e+04	2.612400e+04
max	456191.000000	436802.000000	793319.000000	1.481305e+06	3.011543e+06

#dropping duplicates and nulls

df=df.dropna()

df=df.drop_duplicates()

result = df[df['title'].str.startswith('Harry Potter')]

result.describe()

	book_id	goodreads_book_id	best_book_id	work_id	\
count	9.000000	9.000000	9.000000	9.000000e+00	
mean	479.444444	112688.888889	112688.888889	9.760957e+06	
std	1234.783292	284526.395064	284526.395064	1.328732e+07	
min	2.000000	1.000000	1.000000	2.402163e+06	
25%	21.000000	3.000000	3.000000	2.962492e+06	
50%	24.000000	6.000000	6.000000	3.046572e+06	
75%	27.000000	15881.000000	15881.000000	6.231171e+06	
max	3753.000000	862041.000000	862041.000000	4.133543e+07	

	books_count	isbn13	original_publication_year
average_rating	\		
count	9.000000	9.000000e+00	9.000000
	9.0000		

mean	280.444444	9.780463e+12	2001.333333
4.5500			
std	153.337138	4.654103e+07	3.708099
0.1253			
min	6.000000	9.780439e+12	1997.000000
4.3700			
25%	263.000000	9.780439e+12	1998.000000
4.4600			
50%	307.000000	9.780440e+12	2000.000000
4.5300			
75%	376.000000	9.780440e+12	2005.000000
4.6100			
max	491.000000	9.780545e+12	2007.000000
4.7400			

	ratings_count	work_ratings_count	work_text_reviews_count	\
count	9.000000e+00	9.000000e+00	9.000000	
mean	1.704790e+06	1.805367e+06	32528.777778	
std	1.301040e+06	1.357471e+06	22335.707549	
min	2.461800e+04	2.627400e+04	882.000000	
25%	1.678823e+06	1.785676e+06	27520.000000	
50%	1.746574e+06	1.847395e+06	31084.000000	
75%	1.779331e+06	1.906199e+06	36099.000000	
max	4.602479e+06	4.800065e+06	75867.000000	

	ratings_1	ratings_2	ratings_3	ratings_4	ratings_5
count	9.000000	9.000000	9.000000	9.000000e+00	9.000000e+00
mean	13850.666667	29039.888889	161493.111111	4.534314e+05	1.147551e+06
std	23359.121891	30238.686638	134964.423859	3.345266e+05	8.488109e+05
min	203.000000	186.000000	946.000000	3.891000e+03	2.104800e+04
25%	6676.000000	20210.000000	113646.000000	3.839140e+05	1.065084e+06
50%	7308.000000	21516.000000	151785.000000	4.944270e+05	1.161491e+06
75%	9363.000000	31577.000000	180210.000000	5.094470e+05	1.266670e+06
max	75504.000000	101676.000000	455024.000000	1.156318e+06	3.011543e+06

```
average_rating=result['average_rating'].mean()
average_rating
```

```
4.5500000000000001
```

```
maximum_books_count=result['books_count'].max()
max_sold_book=result.loc[result['books_count'].idxmax(),'title']
print('max sold book is',max_sold_book,':',maximum_books_count)
```

```
max sold book is Harry Potter and the Sorcerer's Stone (Harry Potter,
#1) : 491
```

```
import nbconvert as nb
jupyter nbconvert --to books_analysis.ipynb
```

```
Cell In[3], line 2
```

```
    jupyter nbconvert --to books_analysis.ipynb
    ^
```

```
SyntaxError: invalid syntax
```