

```
[4]: import pandas as pd

df = pd.read_csv("books.csv")

print(df.head())
print(df.info())
null_values = df.isnull().sum()

print("Null values in each column: ")
print(null_values)

if df.duplicated().any():
    print("There are duplicate rows in the DataFrame. ")
else:
    print("There are no duplicate rows in the DataFrame. ")

df_cleaned = df.dropna()

output_path = 'cleaned_books.csv'

df_cleaned.to_csv(output_path, index=False)

print("Cleaned data saved to: ", output_path)

df_clean = pd.read_csv('cleaned_books.csv')
null_values = df_clean.isnull().sum()

print("Null values in each column: ")
print(null_values)

new_df = pd.read_csv('cleaned_books.csv')

harry_potter_df = new_df[new_df[original_title].str.contains('Harry Potter', _
    ↵case=False)]

harry_potter_df.to_csv('Harry_Potter_Books.csv', index=False)
```

```

harry_df = pd.read_csv(r'Harry_Potter_Books.csv')

sorted_books = harry_df.sort_values(by=['ratings_count', 'average_rating'],
                                     ascending=[False, False])

most_selling = sorted_books.iloc[0]

print("The Most selling book based on average rating and ratings count: ")
print(most_selling)

average_rating = harry_potter_df[average_rating].mean()

print("Average rating:", average_rating)

```

	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]

<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

```

```

None

```

```

Null values in each

```

```

column:

```

```

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

```

```

There are no duplicate rows in the

```

```

DataFrame. Cleaned data saved to:

```

```

cleaned_books.csv Null values in each column:

```

```

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 The Most selling book based on average rating and
ratings count: book_id 2 goodreads_book_id 3 best_book_id 3 work_id
4640799 books_count 491 isbn 439554934 isbn13 9780439554930.0
authors      J.K. Rowling, Mary GrandPré
original_publication_year 1997.0 original_title Harry Potter and
the Philosopher's Stone title Harry Potter and the Sorcerer's Stone
(Harry P..
language_code      eng average_rating 4.44 ratings_count
4602479 work_ratings_count 4800065 work_text_reviews_count 75867
ratings_1 75504 ratings_2 101676 ratings_3 455024 ratings_4
1156318 ratings_5      3011543
image_url          https://images.gr-assets.com/books/1474154022m...
small_image_url    https://images.gr-assets.com/books/1474154022s...
Name: 0, dtype: object
Average rating: 4.4910000000000005

```

```
[ ]:
```

