

```
In [72]: import pandas as pd
```

Load and display the dataset

```
In [73]: df = pd.read_csv("D:\\assignment cloud\\books.csv")
```

```
print("Original Dataset:")
print(df.head())
```

Original Dataset:

	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]

Check for missing values

```
In [74]: print("\nMissing Values:")
print(df.isnull().sum())
```

```
Missing Values:
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
```

```
In [75]: # Check for duplicates
print("\nDuplicate Rows:")
print(df[df.duplicated()])
```

Duplicate Rows:
Empty DataFrame

Columns: [book_id, goodreads_book_id, best_book_id, work_id, books_count, isbn, isbn13, authors, original_publication_year, original_title, title, language_code, average_rating, ratings_count, work_ratings_count, work_text_reviews_count, ratings_1, ratings_2, ratings_3, ratings_4, ratings_5, image_url, small_image_url]
Index: []

[0 rows x 23 columns]

Drop duplicates

```
In [76]: df.drop_duplicates(inplace=True)
```

Display information about the Harry Potter subset

```
In [77]: # Subset creation: Extracting only the Harry Potter books
        harry_potter_df = df[df['title'].str.contains("Harry Potter")]

        print("\nHarry Potter Subset:")
        print(harry_potter_df)
        # Save the cleaned dataset to a new CSV file
        harry_potter_df.to_csv("harry_potter_books.csv", index=False)
```

```

Harry Potter Subset:
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
1266      9048              2002       2002     8621948       5

      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
1266 043932162X  9.780439e+12      J.K. Rowling

      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
1266       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
1266                                     NaN  ...      10736

      work_ratings_count  work_text_reviews_count  ratings_1  ratings_2  \
1          4800065          75867          7504      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
1266         11732           185          106      304

      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
1266         1548       2595       7179

      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...
1266 https://images.gr-assets.com/books/1479775271m...

      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...
1266 https://images.gr-assets.com/books/1479775271s...

```

[11 rows x 23 columns]

Data Cleaning and Preprocessing

```
In [78]: # Subset Creation for Harry Potter books
harry_potter_df = df[df['title'].str.contains("Harry Potter")]

# Find the most selling books within the Harry Potter series
most_selling_books = harry_potter_df.sort_values(by='ratings_count', ascending=False).head()

# Calculate the average rating of the Harry Potter books
average_rating = harry_potter_df['average_rating'].mean()

# Display the most selling books and average rating
print("\nMost Selling Books within the Harry Potter Series:")
print(most_selling_books[['title', 'ratings_count']])
print("\nAverage Rating of the Harry Potter Books:", average_rating)
```

```
Most Selling Books within the Harry Potter Series:
               title  ratings_count
1  Harry Potter and the Sorcerer's Stone (Harry P...    4602479
6  Harry Potter and the Prisoner of Azkaban (Harr...    1832823
9  Harry Potter and the Chamber of Secrets (Harry...    1779331
10 Harry Potter and the Goblet of Fire (Harry Pot...    1753043
11 Harry Potter and the Deathly Hallows (Harry Po...    1746574
```

```
Average Rating of the Harry Potter Books: 4.4827272727273
```