1. Load the data set

1.Load the data set

```
import pandas as pd
# Load the dataset
df = pd.read_csv('books.csv')
print(df.columns.tolist())
```

Python

```
['book_id', 'goodreads_book_id', 'best_book_id', 'work_id', 'books_count', 'isbn', 'isbn13', 'authors', 'original_publication_year', 'original_title', 'title', 'language_code',
```

2. Data cleaning & data processing

2.1. remove rows with missing values.

2.Data cleaning & Data preprocessing

```
#1.Remove rows with missing values:
df.dropna(subset=['book_id', 'original_title', 'title', 'average_rating', 'ratings_count'], inplace=True)
df
```

Python

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings_count	work_ratings_count	w
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	...	4780653	4942365	
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	4602479	4800065	
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	...	3866839	3916824	
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	...	2346404	2478609	
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	...	1903563	2216814	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1349	9925	86737	86737	3877968	52	1582349177	9.781582e+12	Mary Hoffman	2002.0	City of Masks	...	12048	13385	
1350	9937	13010211	13010211	18171867	22	1596435712	9.781596e+12	Caragh M. O'Brien	2012.0	Promised	...	11766	12884	

2.2. Convert ‘ratings_count’ to numeric (replace non-numeric values with NaN):

```
#2.Convert ‘ratings_count’ to numeric (replace non-numeric values with NaN):
df['ratings_count'] = pd.to_numeric(df['ratings_count'], errors='coerce')
df.dropna(subset=['ratings_count'], inplace=True)
print(df['ratings_count'])
```

Python

```
0       4780653
1       4602479
2       3866839
3       2346404
4       1903563
         ...
1349      12048
1350      11766
1351      10439
1352      12510
1353      13954
Name: ratings_count, Length: 1153, dtype: int64
```

2.3. Convert ‘Rating’ to numeric:

```
#3.Convert ‘Rating’ to numeric:
df['average_rating'] = pd.to_numeric(df['average_rating'], errors='coerce')
print(df['average_rating'])
```

Python

```
0       4.34
1       4.44
2       3.57
3       4.26
4       4.24
        ...
1349    3.90
1350    3.77
1351    4.14
1352    3.60
1353    3.95
Name: average_rating, Length: 1153, dtype: float64
```

3.Filter for Harry Potter Series:

```
harry_potter_books = df[df['original_title'].str.contains("Harry Potter", case=False, na=False)]
harry_potter_books
```

Python

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings_count	work_ratings_count	w
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	4602479	4800065	
6	18	5	5	2402163	376	043965548X	9.780440e+12	J.K. Rowling, Mary GrandPré, Rufus Beck	1999.0	Harry Potter and the Prisoner of Azkaban	...	1832823	1969375	
8	21	2	2	2809203	307	439358078	9.780439e+12	J.K. Rowling, Mary GrandPré	2003.0	Harry Potter and the Order of the Phoenix	...	1735368	1840548	
9	23	15881	15881	6231171	398	439064864	9.780439e+12	J.K. Rowling, Mary GrandPré	1998.0	Harry Potter and the Chamber of Secrets	...	1779331	1906199	
10	24	6	6	3046572	332	439139600	9.780439e+12	J.K. Rowling, Mary GrandPré	2000.0	Harry Potter and the Goblet of Fire	...	1753043	1868642	
11	25	136251	136251	2963218	263	545010225	9.780545e+12	J.K. Rowling, Mary	2007.0	Harry Potter and the Deathly	...	1746574	1847395	

11	25	136251	136251	2963218	263	545010225	9.780545e+12	J.K. Rowling, Mary GrandPré	2007.0	Harry Potter and the Deathly Hallows	...	1746574	1847395	
12	27	1	1	41335427	275	439785960	9.780440e+12	J.K. Rowling, Mary GrandPré	2005.0	Harry Potter and the Half-Blood Prince	...	1678823	1785676	
96	422	862041	862041	2962492	76	545044251	9.780545e+12	J.K. Rowling	1998.0	Complete Harry Potter Boxed Set	...	190050	204125	
613	3753	10	10	21457570	6	439827604	9.780440e+12	J.K. Rowling	2005.0	Harry Potter Collection (Harry Potter, #1-6)	...	24618	26274	
1036	7018	483445	483445	471792	42	042519891X	9.780425e+12	David Colbert	2001.0	The Magical Worlds of Harry Potter: A Treasury...	...	13820	15145	

10 rows × 23 columns

4.Find the Most selling Harry Potter Book:

```
most_selling_hp = harry_potter_books.nlargest(10, 'ratings_count')
most_selling_hp
```

Python

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings_count	work_ratings_count	w
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	4602479	4800065	
6	18	5	5	2402163	376	043965548X	9.780440e+12	J.K. Rowling, Mary GrandPré, Rufus Beck	1999.0	Harry Potter and the Prisoner of Azkaban	...	1832823	1969375	
9	23	15881	15881	6231171	398	439064864	9.780439e+12	J.K. Rowling, Mary GrandPré	1998.0	Harry Potter and the Chamber of Secrets	...	1779331	1906199	
10	24	6	6	3046572	332	439139600	9.780439e+12	J.K. Rowling, Mary GrandPré	2000.0	Harry Potter and the Goblet of Fire	...	1753043	1868642	
11	25	136251	136251	2963218	263	545010225	9.780545e+12	J.K. Rowling, Mary GrandPré	2007.0	Harry Potter and the Deathly Hallows	...	1746574	1847395	

4. the rest of the table

11	25	136251	136251	2963218	263	545010225	9.780545e+12	J.K. Rowling, Mary GrandPré	2007.0	Harry Potter and the Deathly Hallows	...	1746574	1847395
8	21	2	2	2809203	307	439358078	9.780439e+12	J.K. Rowling, Mary GrandPré	2003.0	Harry Potter and the Order of the Phoenix	...	1735368	1840548
12	27	1	1	41335427	275	439785960	9.780440e+12	J.K. Rowling, Mary GrandPré	2005.0	Harry Potter and the Half-Blood Prince	...	1678823	1785676
96	422	862041	862041	2962492	76	545044251	9.780545e+12	J.K. Rowling	1998.0	Complete Harry Potter Boxed Set	...	190050	204125
613	3753	10	10	21457570	6	439827604	9.780440e+12	J.K. Rowling	2005.0	Harry Potter Collection (Harry Potter, #1-6)	...	24618	26274
1036	7018	483445	483445	471792	42	042519891X	9.780425e+12	David Colbert	2001.0	The Magical Worlds of Harry Potter: A Treasury...	...	13820	15145

10 rows × 23 columns

5.Calculate the Average Rating of Harry Potter Books:

5.Calculate the Average Rating of Harry Potter Books:

```python
average_rating_hp = harry_potter_books['average_rating'].mean()
print(average_rating_hp)
```

Python

```
4.491000000000005
```

6.Display the Results:

```python
print("Most selling Harry Potter book:")
print(most_selling_hp[['original_title', 'average_rating']])
print(f"\nAverage rating of Harry Potter books: {average_rating_hp:.2f}")
```

Python

```
Most selling Harry Potter book:
                                  original_title  average_rating
1              Harry Potter and the Philosopher's Stone           4.44
6              Harry Potter and the Prisoner of Azkaban           4.53
9               Harry Potter and the Chamber of Secrets           4.37
10                  Harry Potter and the Goblet of Fire           4.53
11                Harry Potter and the Deathly Hallows           4.61
8              Harry Potter and the Order of the Phoenix           4.46
12               Harry Potter and the Half-Blood Prince           4.54
96                     Complete Harry Potter Boxed Set           4.74
613         Harry Potter Collection (Harry Potter, #1-6)           4.73
1036  The Magical Worlds of Harry Potter: A Treasury...           3.96

Average rating of Harry Potter books: 4.49
```

7.Save the Cleaned and Processed Dataset:

7.Save the Cleaned and Processed Dataset:

markdown

```python
df.to_csv('cleaned_books.csv', index=False)
print("\nCleaned dataset saved as 'cleaned_books.csv'.")
```

Python

```
Cleaned dataset saved as 'cleaned_books.csv'.
```