

Aliaa Omar

2206187

127.0.0.1:8888/notebooks/HarryPotter.ipynb

jupyter HarryPotter

File Edit View Run Kernel Settings Help

Not Trusted

Open in... Python 3 (ipykernel)

Importing Libraries used

```
[1]: import numpy as np
import pandas as pd
```

Reading Data

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

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings_count
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	...	478065
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	460247
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	...	386683
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	...	234640
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	...	190356
...

jupyter HarryPotter

File Edit View Run Kernel Settings Help

Not Trusted

Open in... Python 3 (ipykernel)

4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	...	190356
...
1349	9925	86737	86737	3877968	52	1582349177	9.781582e+12	Mary Hoffman	2002.0	City of Masks	...	1204
1350	9937	13010211	13010211	18171867	22	1596435712	9.781596e+12	Caragh M. O'Brien	2012.0	Promised	...	1176
1351	9942	16074758	16074758	21869436	18	1442486597	9.781442e+12	Abigail Haas, Abby McDonald	2013.0	Dangerous Girls	...	1043
1352	9947	21393526	21393526	40690062	19	62320521	9.780062e+12	Maria Dahvana Headley	2015.0	Magonia	...	1251
1353	9955	13065327	13065327	18230950	25	802734375	9.780803e+12	Simone Elkeles	2013.0	Wild Cards	...	1395

1354 rows x 23 columns

Check for missing values

```
[3]: missing_values = df.isna()
print(missing_values)
Cleansdf = df.dropna()
```

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn
0	False	False	False	False	False	False
1	False	False	False	False	False	False

jupyterHarryPotter

File Edit View Run Kernel Settings Help

Not Trusted

Open in...Python 3 (ipykernel)

Check for missing values

```
[3]: missing_values = df.isna()
print(missing_values)
Cleansdf = df.dropna()
```

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...
1349	False	False	False	False	False	False
1350	False	False	False	False	False	False
1351	False	False	False	False	False	False
1352	False	False	False	False	False	False
1353	False	False	False	False	False	False

	isbn13	authors	original_publication_year	original_title	...
0	False	False	False	False	...
1	False	False	False	False	...
2	False	False	False	False	...
3	False	False	False	False	...
4	False	False	False	False	...
...
1349	False	False	False	False	...
1350	False	False	False	False	...
1351	False	False	False	False	...
1352	False	False	False	False	...
1353	False	False	False	False	...

	ratings_count	work_ratings_count	work_text_reviews_count	ratings_1
0	False	False	False	False
1	False	False	False	False
...

jupyterHarryPotter

File Edit View Run Kernel Settings Help

Not Trusted

Open in...Python 3 (ipykernel)

```
...
1349  False  False  False  False  False  False
1350  False  False  False  False  False  False
1351  False  False  False  False  False  False
1352  False  False  False  False  False  False
1353  False  False  False  False  False  False

ratings_2 ratings_3 ratings_4 ratings_5 image_url small_image_url
0  False  False  False  False  False  False
1  False  False  False  False  False  False
2  False  False  False  False  False  False
3  False  False  False  False  False  False
4  False  False  False  False  False  False
...  ...  ...  ...  ...  ...
1349  False  False  False  False  False  False
1350  False  False  False  False  False  False
1351  False  False  False  False  False  False
1352  False  False  False  False  False  False
1353  False  False  False  False  False  False

[1354 rows x 23 columns]
```

```
[4]: harry_potter_books = Cleansdf[Cleansdf["title"].str.contains("Harry Potter", case=False, na=False)]
print(harry_potter_books)
```

	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

```
jupyter HarryPotter
File Edit View Run Kernel Settings Help
Not Trusted
Python 3 (ipykernel)

1036 7018 483445 471792 42

    isbn isbn13 authors \
1 439554934 9.780440e+12 J.K. Rowling, Mary GrandPré 4602479
6 04395548X 9.780440e+12 J.K. Rowling, Mary GrandPré, Rufus Beck 1832823
8 439358078 9.780439e+12 J.K. Rowling, Mary GrandPré 1735368
9 439064864 9.780439e+12 J.K. Rowling, Mary GrandPré 1779331
10 439139600 9.780439e+12 J.K. Rowling, Mary GrandPré 1753043
11 545010225 9.780545e+12 J.K. Rowling, Mary GrandPré 1746574
12 439785960 9.780440e+12 J.K. Rowling, Mary GrandPré 1678823
96 545044251 9.780545e+12 J.K. Rowling 190050
613 439827604 9.780440e+12 J.K. Rowling 24618
1036 042519891X 9.780425e+12 David Colbert 13820

    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

    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
```

```
jupyter HarryPotter
File Edit View Run Kernel Settings Help
Not Trusted
Python 3 (ipykernel)

1036 The Magical Worlds of Harry Potter: A Treasury... 13820

    work_ratings_count work_text_reviews_count ratings_1 ratings_2 \
1 4808065 75867 75584 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 1195065
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/1474154802m...
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...
```

jupyter

HarryPotter

File Edit View Run Kernel Settings Help

Not Trusted

1036 https://s.gr-assets.com/assets/nophoto/book/11...

Code

Open in... Python 3 (ipykernel)

```
1 small_image_url
6 https://images.gr-assets.com/books/1474154822s...
8 https://images.gr-assets.com/books/1499277281s...
9 https://images.gr-assets.com/books/1387141547s...
10 https://images.gr-assets.com/books/1474169725s...
11 https://images.gr-assets.com/books/1361482611s...
12 https://images.gr-assets.com/books/1474171184s...
96 https://images.gr-assets.com/books/1361039191s...
613 https://images.gr-assets.com/books/1392579059s...
1036 https://s.gr-assets.com/assets/nophoto/book/50...
```

[10 rows x 23 columns]

[5]:

```
cdf = harry_potter_books[
    [
        "book_id",
        "best_book_id",
        "books_count",
        "title",
        "average_rating",
    ]
]
cdf.head(10)
```

[5]:

	book_id	best_book_id	books_count	title	average_rating
1	2	3	491	Harry Potter and the Sorcerer's Stone (Harry P...	4.44
6	18	5	376	Harry Potter and the Prisoner of Azkaban (Harr...	4.53
8	21	2	307	Harry Potter and the Order of the Phoenix (Har...	4.46
9	23	15881	398	Harry Potter and the Chamber of Secrets (Harry...	4.37
10	24	6	332	Harrrv Potter and the Goblet of Fire (Harrrv Pot...	4.53

jupyter

HarryPotter

File Edit View Run Kernel Settings Help

Not Trusted

[10 rows x 23 columns]

Code

Open in... Python 3 (ipykernel)

[5]:

```
cdf = harry_potter_books[
    [
        "book_id",
        "best_book_id",
        "books_count",
        "title",
        "average_rating",
    ]
]
cdf.head(10)
```

[5]:

	book_id	best_book_id	books_count	title	average_rating
1	2	3	491	Harry Potter and the Sorcerer's Stone (Harry P...	4.44
6	18	5	376	Harry Potter and the Prisoner of Azkaban (Harr...	4.53
8	21	2	307	Harry Potter and the Order of the Phoenix (Har...	4.46
9	23	15881	398	Harry Potter and the Chamber of Secrets (Harry...	4.37
10	24	6	332	Harry Potter and the Goblet of Fire (Harry Pot...	4.53
11	25	136251	263	Harry Potter and the Deathly Hallows (Harry Po...	4.61
12	27	1	275	Harry Potter and the Half-Blood Prince (Harry ...	4.54
96	422	862041	76	Harry Potter Boxset (Harry Potter, #1-7)	4.74
613	3753	10	6	Harry Potter Collection (Harry Potter, #1-6)	4.73
1036	7018	483445	42	The Magical Worlds of Harry Potter: A Treasury...	3.96