

Docker assignment

Click here to ask Blackbox to help you code faster

```
ab = pd.read_csv('books.csv')
ab
```

[2]

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
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
1351	9942	16074758	16074758	21869436	18	1442486597	9.781442e+12	Abigail Haas, Abby McDonald	2013.0	Dangerous Girls	...	10439	12970
1352	9947	21393526	21393526	40690062	19	62320521	9.780062e+12	Maria Dahvana	2015.0	Magonia	...	12510	13652

Click here to ask Blackbox to help you code faster

```
ab.tail()
ab.info()
```

[3]

<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
...
21 image_url 1354 non-null object
22 small_image_url 1354 non-null object

```
...
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
```

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

💡 Click here to ask Blackbox to help you code faster

```
ab['original_publication_year'] = pd.to_datetime(ab['original_publication_year'])
print(ab['original_publication_year'])
```

[6]

```
...
0      1970-01-01 00:00:00.000002008
1      1970-01-01 00:00:00.000001997
2      1970-01-01 00:00:00.000002005
3      1970-01-01 00:00:00.000002012
4      1970-01-01 00:00:00.000002011
...
1349   1970-01-01 00:00:00.000002002
1350   1970-01-01 00:00:00.000002012
1351   1970-01-01 00:00:00.000002013
1352   1970-01-01 00:00:00.000002015
1353   1970-01-01 00:00:00.000002013
Name: original_publication_year, Length: 1354, dtype: datetime64[ns]
```

```
ab = ab[(ab.books_count>=0)&(ab.ratings_count>=0)&(ab.work_ratings_count>=0)&(ab.work_text_reviews_count>=0)&(ab.ratings_1>=0)&(ab.ratings_2>=0)&(ab.ratings_3>=0)&(ab.ratings_4>=0)]
print(ab)
```

[7]

Python

```
...
   book_id  goodreads_book_id  best_book_id  work_id  books_count  \
0         1         2767052      2767052      2792775          272
1         2             3           3         4640799          491
2         3         41865      41865      3212258          226
3         6        11870085      11870085      16827462          226
4        12        13335037      13335037      13155899          210
...
1349    9925          86737      86737      3877968           52
1350    9937        13010211      13010211      18171867           22
1351    9942        16074758      16074758      21869436           18
1352    9947        21393526      21393526      40690062           19
1353    9955        13065327      13065327      18230950           25

   isbn      isbn13      authors  \
0  439023483  9.780439e+12  Suzanne Collins
1  439554934  9.780440e+12  J.K. Rowling, Mary GrandPré
2  316015849  9.780316e+12  Stephanie Meyer
3  525478817  9.780525e+12  John Green
4  62024035  9.780062e+12  Veronica Roth
...
1349 1582349177  9.781582e+12  Mary Hoffman
1350 1596435712  9.781596e+12  Caragh M. O'Brien
1351 1442486597  9.781442e+12  Abigail Haas, Abby McDonald
1352  62320521  9.780062e+12  Maria Dahvana Headley
1353  802734375  9.780803e+12  Simone Elkeles
...
1352  https://images.gr-assets.com/books/1413479866s...
1353  https://images.gr-assets.com/books/1420836321s...
```

[1354 rows x 23 columns]



Click here to ask Blackbox to help you code faster

```
ab = ab.drop_duplicates()
print(ab)
```

[9]

```
...      book_id  goodreads_book_id  best_book_id  work_id  books_count  \
0         1         2767052         2767052    2792775         272
1         2             3             3        4640799         491
2         3         41865         41865        3212258         226
3         6        11870085        11870085    16827462         226
4        12        13335037        13335037    13155899         210
...      ...      ...      ...      ...      ...
1349     9925         86737         86737     3877968          52
1350     9937        13010211        13010211    18171867          22
1351     9942        16074758        16074758    21869436          18
1352     9947        21393526        21393526    40690062          19
1353     9955        13065327        13065327    18230950          25

      isbn      isbn13      authors  \
0  439023483  9.780439e+12  Suzanne Collins
1  439554934  9.780440e+12  J.K. Rowling, Mary GrandPré
2  316015849  9.780316e+12  Stephenie Meyer
3  525478817  9.780525e+12  John Green
4  62024035  9.780062e+12  Veronica Roth
...      ...      ...      ...
1349 1582349177  9.781582e+12  Mary Hoffman
1350 1596435712  9.781596e+12  Caragh M. O'Brien
1351 1442486597  9.781442e+12  Abigail Haas, Abby McDonald
1352  62320521  9.780062e+12  Maria Dahvana Headley
1353  802734375  9.780803e+12  Simone Elkeles
...
1352 https://images.gr-assets.com/books/1413479866s...
1353 https://images.gr-assets.com/books/1420836321s...
```

[1354 rows x 23 columns]

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...



💡 Click here to ask Blackbox to help you code faster
`ab.isnull().sum()`

[10]

```
... 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
```

💡 Click here to ask Blackbox to help you code faster

```
ab.dropna(inplace=True)
```

```
ab.isnull().sum()
```

[11]

```
... 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
```

Click here to ask Blackbox to help you code faster

```
HarryPotter = ab[ab['original_title'].str.contains('Harry Potter', case=False)]
print (HarryPotter)
```

[12]

```
...      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

      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

      original_publication_year  \
...
613  https://images.gr-assets.com/books/1328867351s...
1036 https://s.gr-assets.com/assets/nophoto/book/50...
```

[10 rows x 23 columns]

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

Click here to ask Blackbox to help you code faster

```
MostSellingBooks = HarryPotter.loc[HarryPotter['best_book_id'].idxmax()]
print (MostSellingBooks)
```

[13]

```
...  book_id      422
      goodreads_book_id      862041
      best_book_id      862041
      work_id      2962492
      books_count      76
      isbn      545044251
      isbn13      9780545044260.0
      authors      J.K. Rowling
      original_publication_year      1970-01-01 00:00:00.000001998
      original_title      Complete Harry Potter Boxed Set
      title      Harry Potter Boxset (Harry Potter, #1-7)
      language_code      eng
      average_rating      4.74
      ratings_count      190050
      work_ratings_count      204125
      work_text_reviews_count      6508
      ratings_1      1105
      ratings_2      1285
      ratings_3      7020
      ratings_4      30666
      ratings_5      164049
      image_url      https://images.gr-assets.com/books/1392579059m...
      small_image_url      https://images.gr-assets.com/books/1392579059s...
      Name: 96, dtype: object
```

💡 Click here to ask Blackbox to help you code faster

```
AverageRating = HarryPotter['average_rating'].mean()  
print (AverageRating)
```

[14]

```
... 4.4910000000000005
```