Name :- Om Dattatray Korde

PRN :- 202401120044

DIV :- CS8-2

Roll No :- 30 *

```
import pandas as pd
import numpy as np
df = pd.read_csv('/content/drive/MyDrive/DATASET1/books.csv', on_bad_lines='skip')
```

1. What is the average number of ratings received by books

```
df['ratings_count'].mean()
```

```
np.float64(17942.848062573048)
```

2. Identify the book with the highest average rating.

```
df.loc[df['average_rating'].idxmax(), ['title', 'average_rating']]
```

	624
title	Comoediae 1: Acharenses/Equites/Nubes/Vespae/P...
average_rating	5.0

dtype: object

3. How many unique authors are there in the dataset?

```
df['authors'].nunique()
```

```
6639
```

4. What is the most common language in which books are written?

```
df['language_code'].mode()[0]
```

5. How many books have received more than 10,000 ratings

```
df[df['ratings_count'] > 10000].shape[0]
```

```
1960
```

6. Find the book with the highest number of ratings.

```
df.loc[df['ratings_count'].idxmax(), ['title', 'ratings_count']]
```

	10336
title	Twilight (Twilight #1)
ratings_count	4597666

7. List all books published by 'Scholastic Inc.'

```
df[df['publisher'] == 'Scholastic Inc.']['title']
```

	title
0	Harry Potter and the Half-Blood Prince (Harry ...
1	Harry Potter and the Order of the Phoenix (Har...
3	Harry Potter and the Prisoner of Azkaban (Harr...
2837	The Long-Lost Map (Ulysses Moore #2)
5531	The Mark of the Crown (Star Wars: Jedi Apprent...
5724	Math Fables
5900	Hush! A Thai Lullaby
6189	The Door to Time (Ulysses Moore #1)
6246	The Journal of Scott Pendleton Collins: A Worl...
6746	King of the Middle March
8204	Ruby the Red Fairy (Rainbow Magic #1)
9467	The Eternity Code (Artemis Fowl #3)
11066	The Wish List

8. Compute the average rating for books written in English.

python Copy code

```
df[df['language_code'] == 'eng']['average_rating'].mean()
```

```
np.float64(3.934061517736866)
```

9. Count the number of books that have "Harry Potter" in the title.

python Copy code

```
df[df['title'].str.contains('Harry Potter', case=False, na=False)].shape[0]
```

```
26
```

10. Show top 5 authors with the most books in the dataset.

python Copy code

```
df['authors'].value_counts().head(5)
```

authors	count
Stephen King	40
P.G. Wodehouse	40
Rumiko Takahashi	39
Orson Scott Card	35
Agatha Christie	33

11. How many books have no text reviews?

python Copy code

```
df[df['text_reviews_count'] == 0].shape[0]
```

```
624
```

12. What is the oldest publication year available in the dataset?

python Copy code

```
pd.to_datetime(df['publication_date'], errors='coerce').dt.year.min()
```

```
1900.0
```

13. Create a new column that calculates the ratio of text reviews to total ratings.

python Copy code

```
df['review_ratio'] = df['text_reviews_count'] / df['ratings_count']
```

14. Find the most prolific author (most books).

python Copy code

```
df['authors'].value_counts().idxmax()
```

15. Find the number of books with a rating greater than the average rating of all books.

python Copy code

```
avg_rating = df['average_rating'].mean()
df[df['average_rating'] > avg_rating].shape[0]
```

```
5960
```

16. Which publisher has published the most books?

python Copy code

```
df['publisher'].value_counts().idxmax()
```

17. What is the total number of text reviews in the dataset

```
df['text_reviews_count'].sum()
```

```
np.int64(6029201)
```

18. List all books with more than 1,000 text reviews.

```
df[df['text_reviews_count'] > 1000][['title', 'text_reviews_count']]
```

	title	text_reviews_count
0	Harry Potter and the Half-Blood Prince (Harry ...	27591
1	Harry Potter and the Order of the Phoenix (Har...	29221
3	Harry Potter and the Prisoner of Azkaban (Harr...	36325
8	The Ultimate Hitchhiker's Guide to the Galaxy ...	4080
12	A Short History of Nearly Everything	9396
...	...	...
11021	The Gun Seller	1434
11024	Pandora's Star	1519
11057	Arthur & George	1141
11065	The Supernaturalist	1001
11097	Undaunted Courage: The Pioneering First Missio...	1830

1076 rows × 2 columns

19. What is the correlation between average rating and text review count python Copy code

```
df[['average_rating', 'text_reviews_count']].corr()
```

	average_rating	text_reviews_count
average_rating	1.000000	0.033663
text_reviews_count	0.033663	1.000000

20. Find how many books have average rating exactly equal to 5.

python Copy code

```
df[df['average_rating'] == 5].shape[0]
```

22