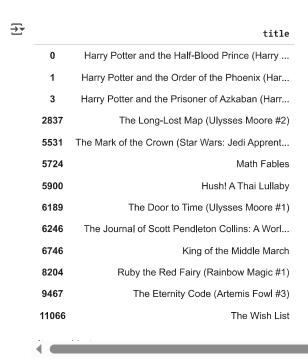
```
Name :- Om Dattatray Korde
PRN: - 202401120044
DIV:- CS8-2
Roll No :- 30 *
import pandas as pd
import numpy as np
df = pd.read_csv('/content/drive/MyDrive/DATASET1/books.csv', on_bad_lines='skip')
   1. What is the average number of ratings received by books
df['ratings_count'].mean()
np.float64(17942.848062573048)
   2. Identify the book with the highest average rating.
df.loc[df['average_rating'].idxmax(), ['title', 'average_rating']]
₹
           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()
\rightarrow
     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']]
\overline{\mathbf{x}}
                                 10336
           title
                     Twilight (Twilight #1)
                               4597666
      ratings_count
```

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

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



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

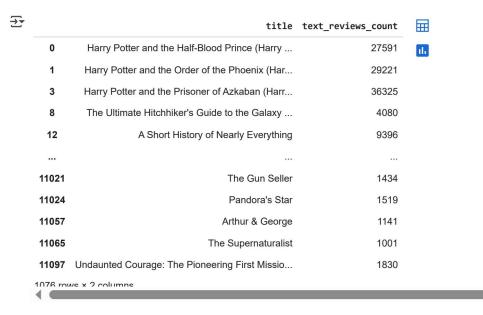
```
df['authors'].value_counts().head(5)
```

count
authors

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]
\rightarrow
    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']]



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

df[['average\_rating', 'text\_reviews\_count']].corr()



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

python Copy code

df[df['average\_rating'] == 5].shape[0]