Name :- Anisha Ghodeswar

PRN :- 202401110001

DIV :- CS7-74

Roll No :- 72

import pandas as pd import numpy as np

df = pd.read\_csv('/content/drive/MyDrive/DATASET1/books.csv', on\_bad\_1ines='skip')

## What is the average number of ratings received by books

df['ratings\_count'].mean()

-t- np.f1oat64(17942.848062573048)

## Identify the book with the highest average rating.

df.loc[df['average\_rating'].idxmax(), ['title', 'average\_rating']]



title

average\_rating

dtype: object

624

Comoediae 1: Acharenses/Equites/Nubes/Vespae/P...

5.0

## How many unique authors are there in the dataset?

d-F[ ' aut hons ' ] . nuni que( )

 6639

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

df['language\_code’].mode()[0]



## How many books have received more than 10,000 ratings

df[df['ratings\_count'] > 10000].shape[0]

 1960

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

df.loc[df['ratings\_count'].idxmax(), ['title', 'ratings\_count']]



title

ratings\_count

10336

Twilight (Twilight #1)

4597666



1. List all books published by 'Scholastic Inc.' df[df['publisher'] == 'Scholastic Inc. ']['title']

title

* 1. Harry Potter and the Half-Blood Prince (Harry ...
  2. 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

**5900**

Math Fables Hush! A Thai Lullaby

6189 The Door to Time (Ulysses Moore #1)



6246 The Journal of Scott Pendleton Collins: A Wort... 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

1. Compute the average rating for books written in English. python Copy code

df[df['language\_code'] == 'eng']['average\_rating'].mean()

-\*\_j np.float64(3.934061517736866)

1. 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

## 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

## How many books have no text reviews? python Copy code

df[df['text\_reviews\_count'] == 0].shape[0]

 624

# J 2. What is the oldest publication year available in the dataset? python Copy code

pd.to\_datetime(df['publication\_date'], errons='coerce').dt.year.min()

\_;• 1900.0

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

## python Copy code

df['review\_ratio'] = df['text\_neviews\_count'] / df['ratings\_count']

## Find the most prolific author (most books). python Copy code

df['authors'].value\_counts().idxmax()



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

## python Copy code

avg\_rating = df['avenage\_rating'].mean() df[df['average\_rating'] > avg\_rating].shape[0]

 5960

# J6. Which publisher has published the most books? python Copy code

df['publisher'].value\_counts().idxmax()



# J7. What is the total number of text reviews in the dataset

df['text\_neviews\_count'].sum()

\* np.int64(6029201)

# J8. 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 |

nyn rn»io x 0 ooh moo



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

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



#### average\_rating

average\_rating text\_reviews\_count :p 1.000000 0.033663



## Find how many books have average rating exactly equal to 5. python Copy code

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

22