

GOODREADS BOOK REVIEWS DATASET ANALYSIS

Roll No: CC 72
PRN: 202401030004
Division: CC

EACH OF THE 10 PROBLEMS IS SOLVED STRICTLY USING PANDAS

1. Find the total number of books.

```
total_books = df.shape[0]
```

2. Find the number of unique authors.

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

3. Find the most common language code.

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

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

```
most_rated_book = df.loc[df['ratings_count'].idxmax()]
```

5. Find the publisher with maximum books.

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

6. Calculate average of average_rating for books published after 2010.

```
df['publication_year'] = pd.to_datetime(df['publication_date'], errors='coerce').dt.year  
avg_rating_post_2010 = df[df['publication_year'] > 2010]['average_rating'].mean()
```

7. Find how many books have more than 5000 text reviews.

```
high_text_review_books = df[df['text_reviews_count'] > 5000].shape[0]
```

8. List all books with missing number of pages.

```
Analysis missing_pages = df[df['num_pages'].isna()]
```

9. Create a column ratings_per_page and find the maximum value.

```
df['ratings_per_page'] = df['ratings_count'] / df['num_pages']  
max_rpp = df['ratings_per_page'].max()
```

10. Find the median number of pages for English books.

```
median_pages_eng = df[df['language_code'] == 'eng']['num_pages'].median()
```

EACH OF THE 10 PROBLEMS IS SOLVED STRICTLY USING NUMPY

11. Find mean ratings count using Numpy.

```
mean_ratings_count = np.mean(ratings_count)
```

12. Find standard deviation of average ratings.

```
std_avg_rating = np.std(average_rating)
```

13. Find maximum number of pages using Numpy.

```
max_pages = np.max(num_pages)
```

14. Find number of books published before 1990 using Numpy.

```
books_before_1990 = np.sum(years < 1990)
```

15. Find the minimum number of text reviews.

```
min_text_reviews = np.min(text_reviews_count)
```



11. Find mean ratings count using Numpy.

```
mean_ratings_count = np.mean(ratings_count)
```

12. Find standard deviation of average ratings.

```
std_avg_rating = np.std(average_rating)
```

13. Find maximum number of pages using Numpy.

```
max_pages = np.max(num_pages)
```

14. Find number of books published before 1990 using Numpy.

```
books_before_1990 = np.sum(years < 1990)
```

15. Find the minimum number of text reviews.

```
min_text_reviews = np.min(text_reviews_count)
```



Thank You

