assignment

April 23, 2024

[1]:
```
import pandas as pd

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

[2]:
```
print(df.isnull().sum())
df.dropna(inplace=True)
```

```
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
```

[3]:
```
harry_potter_df = df[df['original_title'].str.contains('Harry Potter',␣
↪case=False)]

harry_potter_sales = harry_potter_df.groupby('original_title')['books_count'].
↪sum().reset_index()
```

```
most_selling_books = harry_potter_sales.sort_values(by='books_count',␣
 ↪ascending=False).head(5)

print(most_selling_books)
```

```
                          original_title  books_count
7   Harry Potter and the Philosopher's Stone          491
2    Harry Potter and the Chamber of Secrets          398
8   Harry Potter and the Prisoner of Azkaban          376
4        Harry Potter and the Goblet of Fire          332
6  Harry Potter and the Order of the Phoenix          307
```

[4]:
```
average_rating = harry_potter_df['average_rating'].mean()

print("Average rating of Harry Potter books:", average_rating)
```

```
Average rating of Harry Potter books: 4.4910000000000005
```

[ ]:
```

```