

cc-jupyter-notebook-ipynb

April 24, 2024

```
[1]: #importing dependencies
import pandas as pd

[2]: #importing intended csv file
df = pd.read_csv("books.csv")

[3]: #performing data preprocessing and cleaning
df_cleaned = df.dropna(subset=['original_title'])

df_cleaned = df_cleaned.drop_duplicates(subset=['original_title'])

[7]: #filtering the most selling books out of the Harry Potter Series

harry_potter_df = df_cleaned[df_cleaned['original_title'].str.contains('Harry_
↳Potter', case=False)]

most_selling_books = harry_potter_df.groupby('title')['ratings_count'].max().
↳sort_values(ascending=False).head(5)

#Calculating the Average rating of Harry Potter Series
average_rating = harry_potter_df['average_rating'].mean()

[8]: #Displaying the results
print("Most Selling Books within the Harry Potter Series:")
print(most_selling_books)
print("\nAverage Rating of the Harry Potter Books:", average_rating)
```

Most Selling Books within the Harry Potter Series:

title	
Harry Potter and the Sorcerer's Stone (Harry Potter, #1)	4602479
Harry Potter and the Prisoner of Azkaban (Harry Potter, #3)	1832823
Harry Potter and the Chamber of Secrets (Harry Potter, #2)	1779331
Harry Potter and the Goblet of Fire (Harry Potter, #4)	1753043
Harry Potter and the Deathly Hallows (Harry Potter, #7)	1746574

Name: ratings_count, dtype: int64

Average Rating of the Harry Potter Books: 4.4910000000000005