

## Analysis File on Dataset books.csv

Firstly, I made a cleaning process on dataset to remove null values.

### 1. Check for null values and duplicates:

```
df.isnull().sum()
df.duplicated()

# output null
isbn                52
isbn13              44
original_publication_year    3
original_title       52
language_code       109
dtype: int64
# These all columns that have null values.
# output Duplicates
0      False
1      False
2      False
3      False
4      False
...
1349   False
1350   False
1351   False
1352   False
1353   False
Length: 1153, dtype: bool
```

### 2. Drop null values:

```
df = df.dropna()
```

Secondly, Filtering Harry Potter books from dataset:

```
harry_potter_books = df[df["title"].str.contains("Harry Potter", case = False)]
```

Thirdly, Finding most selling book of Harry Potter:

```
most_selling_books = harry_potter_books.sort_values(by = "ratings_count", ascending=False)  
most_selling_books
```

Finally, Calculating the average rating of the Harry Potter books.

```
averageRating = harry_potter_books["average_rating"].mean()  
print(f"Average Rating of Harry Potter Books Is:{averageRating}")
```

Name: Mohamed Said Mohamed Abdo Mekawy

ID: 2206201