

Name: Yasmin Mohamed Mouawad id:2102102

## the results from the analysis.

### # Basic statistics

```
# Basic statistics
cleaned_ds.describe()
```

|       | books_count | original_publication_year | average_rating | ratings_count | work_ratings_count | work_text_reviews_count | ratings_1     | ratings_2     | ratings_3     |
|-------|-------------|---------------------------|----------------|---------------|--------------------|-------------------------|---------------|---------------|---------------|
| count | 1153.000000 | 1153.000000               | 1153.000000    | 1.153000e+03  | 1.153000e+03       | 1153.000000             | 1153.000000   | 1153.000000   | 1153.000000   |
| mean  | 52.777103   | 2003.698179               | 3.992819       | 1.014850e+05  | 1.097655e+05       | 5621.572420             | 2555.634866   | 5540.813530   | 19310.74935   |
| std   | 64.054893   | 15.992755                 | 0.223829       | 3.095194e+05  | 3.258908e+05       | 11410.001335            | 14830.209633  | 17529.393371  | 46821.01928   |
| min   | 2.000000    | 1871.000000               | 3.230000       | 8.861000e+03  | 1.025100e+04       | 112.000000              | 33.000000     | 133.000000    | 946.000000    |
| 25%   | 24.000000   | 2003.000000               | 3.840000       | 1.858300e+04  | 2.030900e+04       | 1327.000000             | 330.000000    | 1041.000000   | 4333.000000   |
| 50%   | 38.000000   | 2008.000000               | 4.000000       | 3.139500e+04  | 3.398500e+04       | 2441.000000             | 688.000000    | 1884.000000   | 7170.000000   |
| 75%   | 61.000000   | 2011.000000               | 4.150000       | 6.682300e+04  | 7.233900e+04       | 4973.000000             | 1505.000000   | 3945.000000   | 14260.000000  |
| max   | 1314.000000 | 2017.000000               | 4.740000       | 4.780653e+06  | 4.942365e+06       | 155254.000000           | 456191.000000 | 436802.000000 | 793319.000000 |

### #Calculate total ratings count for each book

### # Sort the dataset by total ratings in descending order

```
#Calculate total ratings count for each book
cleaned_ds['total_ratings'] = cleaned_ds['ratings_1'] + cleaned_ds['ratings_2'] + cleaned_ds['ratings_3'] + cleaned_ds['ratings_4'] + cleaned_ds['ratings_5']

# Sort the dataset by total ratings in descending order
most_selling_books = cleaned_ds[['title', 'total_ratings']].sort_values(by='total_ratings', ascending=False)

print("Top 5 most selling books within the Harry Potter series:")
print(most_selling_books.head(20))
```

|    | title   | total_ratings |
|----|---|---------------|
| 0  | The Hunger Games (The Hunger Games, #1)           | 4942365       |
| 1  | Harry Potter and the Sorcerer's Stone (Harry P... | 4800065       |
| 2  | Twilight (Twilight, #1)                           | 3916824       |
| 3  | The Fault in Our Stars                            | 2478609       |
| 4  | Divergent (Divergent, #1)                         | 2216814       |
| 5  | Catching Fire (The Hunger Games, #2)              | 1988079       |
| 6  | Harry Potter and the Prisoner of Azkaban (Harr... | 1969375       |
| 7  | Harry Potter and the Chamber of Secrets (Harry... | 1906199       |
| 8  | Mockingjay (The Hunger Games, #3)                 | 1870748       |
| 9  | Harry Potter and the Goblet of Fire (Harry Pot... | 1868642       |
| 10 | Harry Potter and the Deathly Hallows (Harry Po... | 1847395       |
| 11 | Harry Potter and the Order of the Phoenix (Har... | 1840548       |
| 12 | Harry Potter and the Half-Blood Prince (Harry ... | 1785676       |
| 13 | The Lion, the Witch, and the Wardrobe (Chronic... | 1584884       |
| 14 | The Lightning Thief (Percy Jackson and the Oly... | 1411114       |
| 15 | The Book Thief                                    | 1287798       |
| 16 | City of Bones (The Mortal Instruments, #1)        | 1241799       |
| 17 | New Moon (Twilight, #2)                           | 1199000       |
| 18 | Eclipse (Twilight, #3)                            | 1176642       |
| 19 | Eragon (The Inheritance Cycle, #1)                | 1125231       |

Top 5 most selling books within the Harry Potter series:

|   | title   | total_ratings |
|---|---|---------------|
| 0 | The Hunger Games (The Hunger Games, #1)           | 4942365       |
| 1 | Harry Potter and the Sorcerer's Stone (Harry P... | 4800065       |
| 2 | Twilight (Twilight, #1)                           | 3916824       |
| 3 | The Fault in Our Stars                            | 2478609       |
| 4 | Divergent (Divergent, #1)                         | 2216814       |
| 5 | Catching Fire (The Hunger Games, #2)              | 1988079       |

## # Calculate the average rating of the Harry Potter books

```
# Calculate the average rating of the Harry Potter books
average_rating = cleaned_ds['average_rating'].mean()
print("\nAverage rating of the Harry Potter books:", average_rating)
```

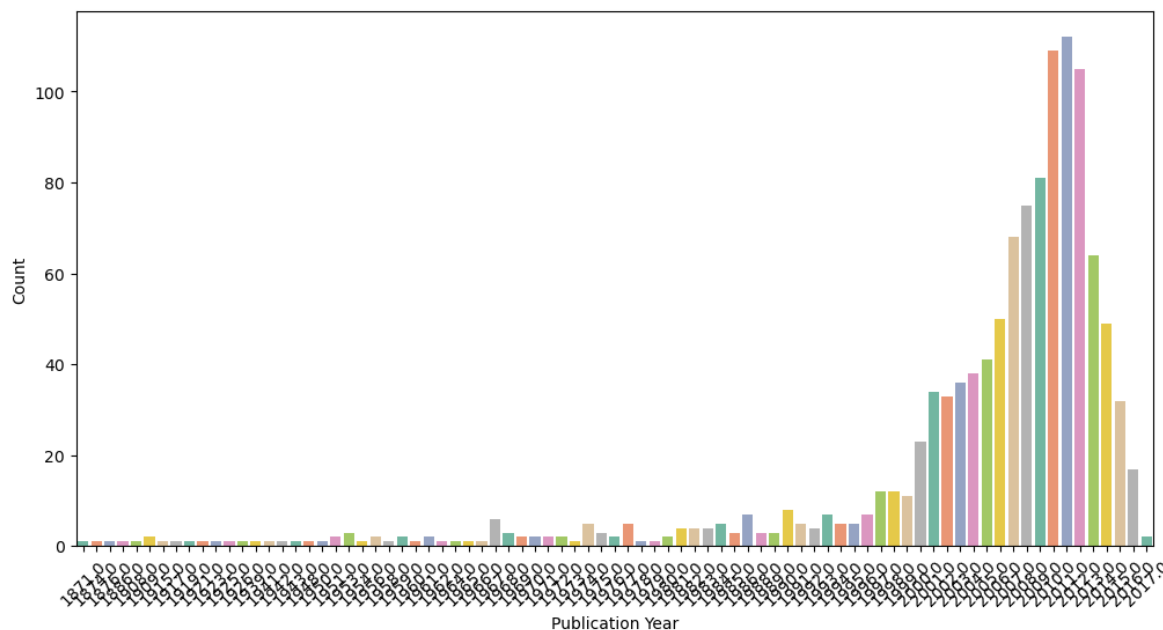
Average rating of the Harry Potter books: 3.992818733738074

## # Distribution of publication years

```
plt.figure(figsize=(12, 6))

sns.countplot(x='original_publication_year', hue='original_publication_year',
data=cleaned_ds, palette='Set2', legend=False)

plt.xlabel('Publication Year')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



## # Ratings count of the Harry Potter books

```
[23]: # Ratings count of the Harry Potter books
ratings_count_hp = cleaned_ds[cleaned_ds['title'].str.contains('Harry Potter')]['ratings_count'].sum()
print("Total ratings count of the Harry Potter books:", ratings_count_hp)
```

Total ratings count of the Harry Potter books: 15356929

