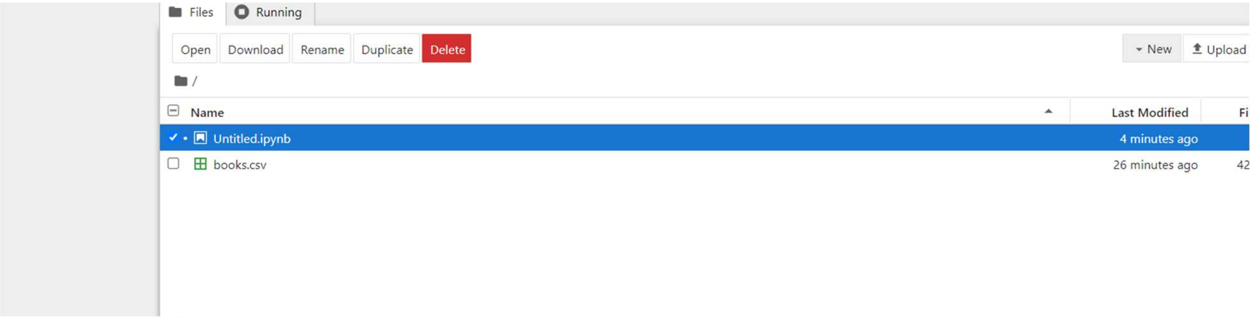
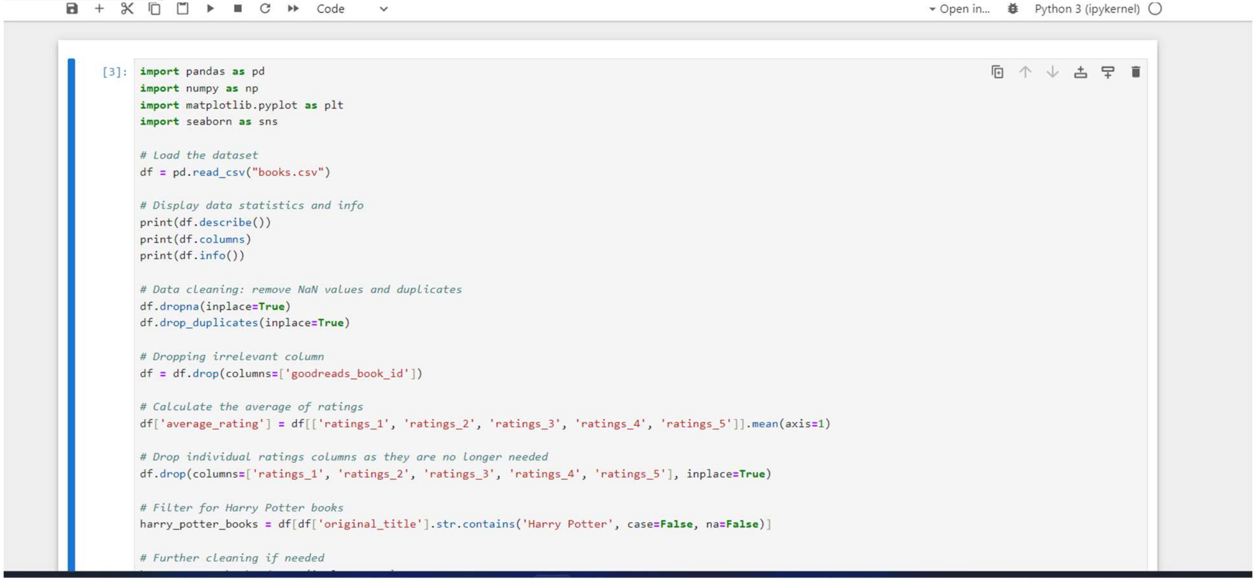
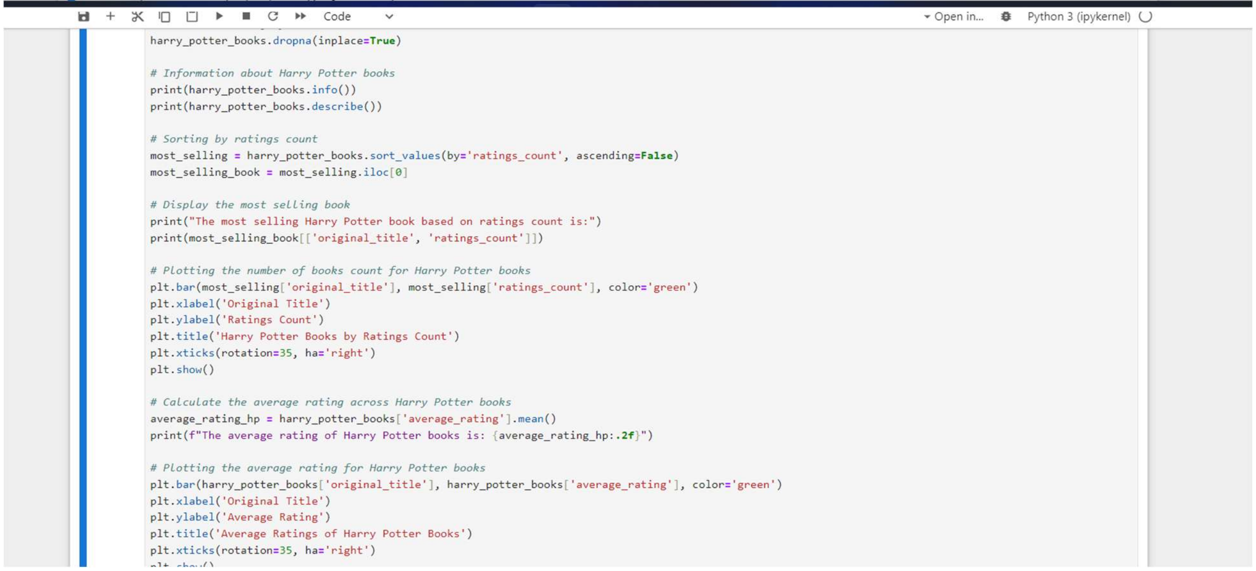
Marwan abdelhamid   2206174



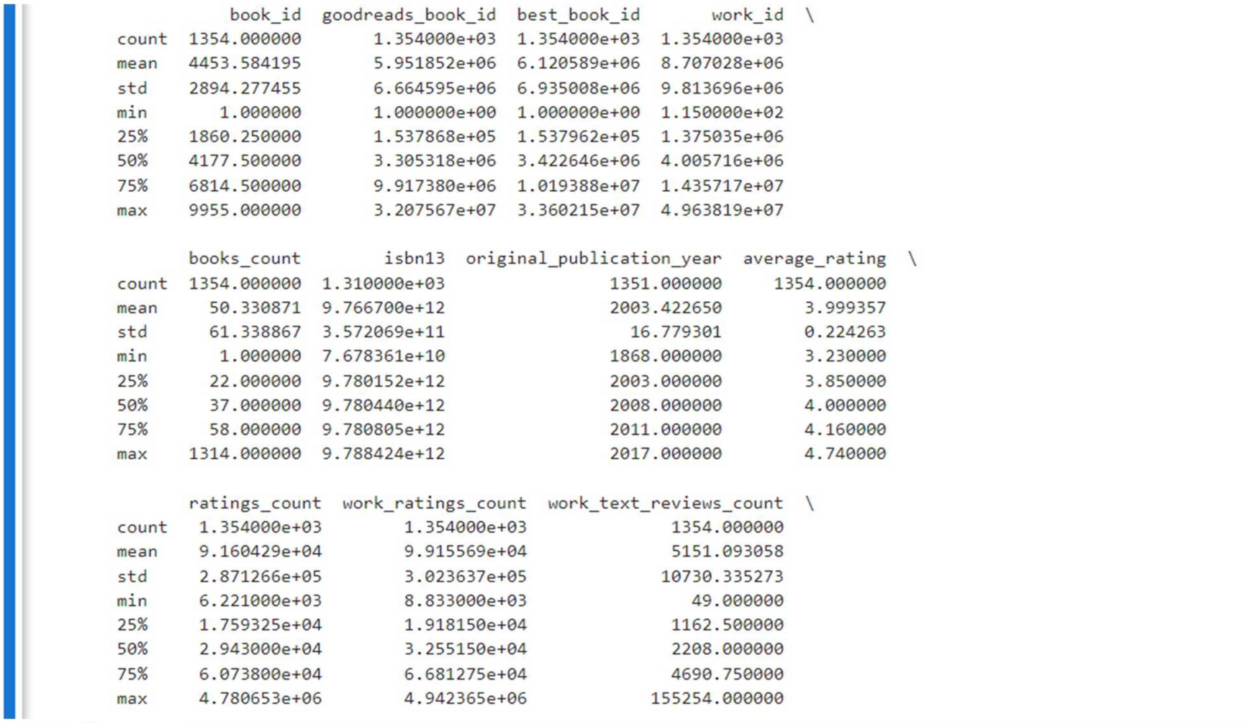
Name	Last Modified	Fi
☑ Untitled.ipynb	4 minutes ago	
□ books.csv	26 minutes ago	42



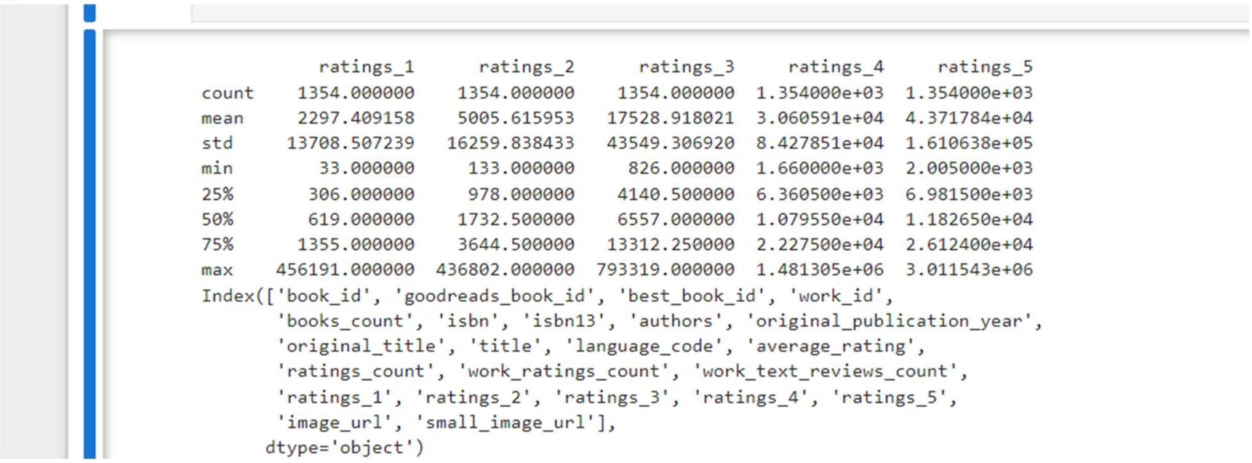
```
[3]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns

     # Load the dataset
     df = pd.read_csv("books.csv")

     # Display data statistics and info
     print(df.describe())
     print(df.columns)
     print(df.info())

     # Data cleaning: remove NaN values and duplicates
     df.dropna(inplace=True)
     df.drop_duplicates(inplace=True)

     # Dropping irrelevant column
     df = df.drop(columns=['goodreads_book_id'])

     # Calculate the average of ratings
     df['average_rating'] = df[['ratings_1', 'ratings_2', 'ratings_3', 'ratings_4', 'ratings_5']].mean(axis=1)

     # Drop individual ratings columns as they are no longer needed
     df.drop(columns=['ratings_1', 'ratings_2', 'ratings_3', 'ratings_4', 'ratings_5'], inplace=True)

     # Filter for Harry Potter books
     harry_potter_books = df[df['original_title'].str.contains('Harry Potter', case=False, na=False)]

     # Further cleaning if needed
```



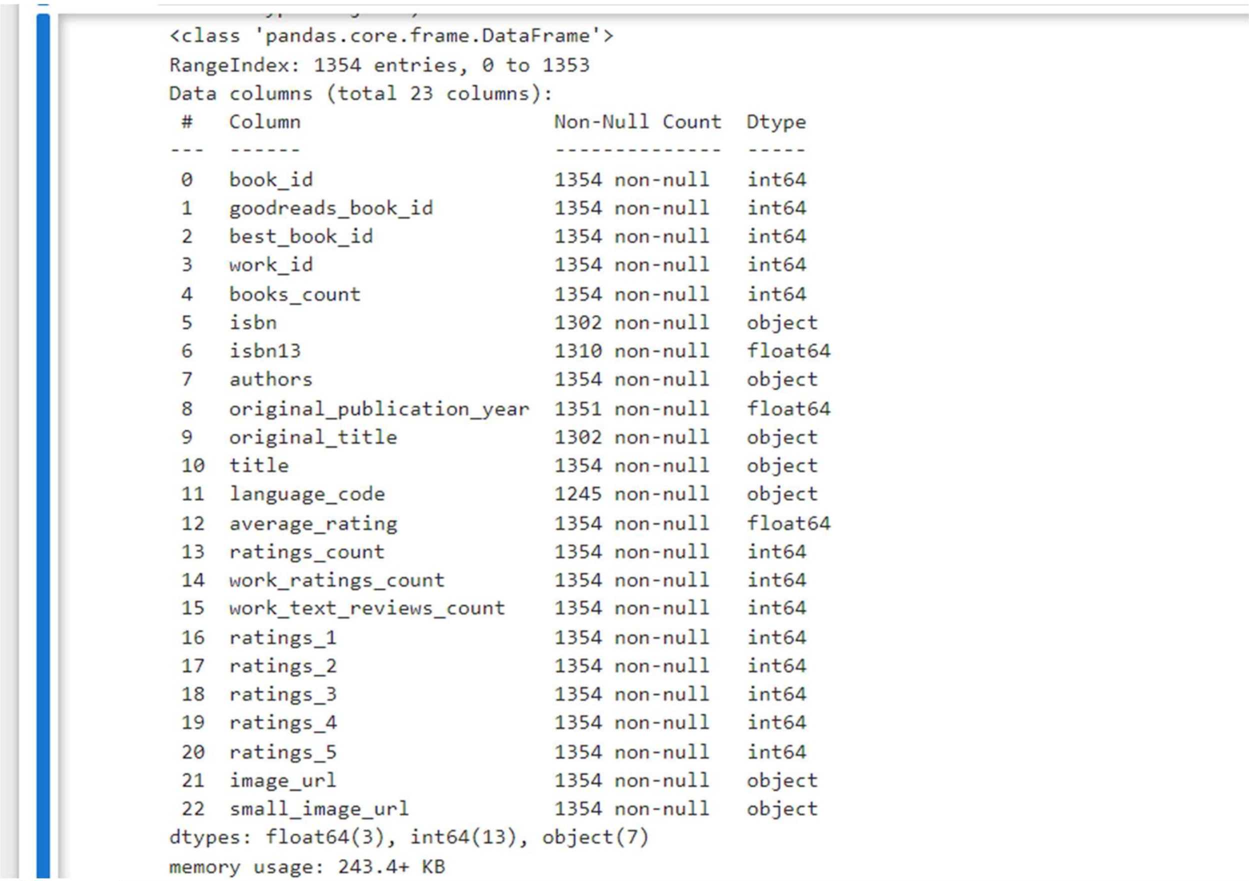
```
harry_potter_books.dropna(inplace=True)

# Information about Harry Potter books
print(harry_potter_books.info())
print(harry_potter_books.describe())

# Sorting by ratings count
most_selling = harry_potter_books.sort_values(by='ratings_count', ascending=False)
most_selling_book = most_selling.iloc[0]

# Display the most selling book
print("The most selling Harry Potter book based on ratings count is:")
print(most_selling_book[['original_title', 'ratings_count']])

# Plotting the number of books count for Harry Potter books
plt.bar(most_selling['original_title'], most_selling['ratings_count'], color='green')
plt.xlabel('Original Title')
plt.ylabel('Ratings Count')
plt.title('Harry Potter Books by Ratings Count')
plt.xticks(rotation=35, ha='right')
plt.show()

# Calculate the average rating across Harry Potter books
average_rating_hp = harry_potter_books['average_rating'].mean()
print(f"The average rating of Harry Potter books is: {average_rating_hp:.2f}")

# Plotting the average rating for Harry Potter books
plt.bar(harry_potter_books['original_title'], harry_potter_books['average_rating'], color='green')
plt.xlabel('Original Title')
plt.ylabel('Average Rating')
plt.title('Average Ratings of Harry Potter Books')
plt.xticks(rotation=35, ha='right')
```



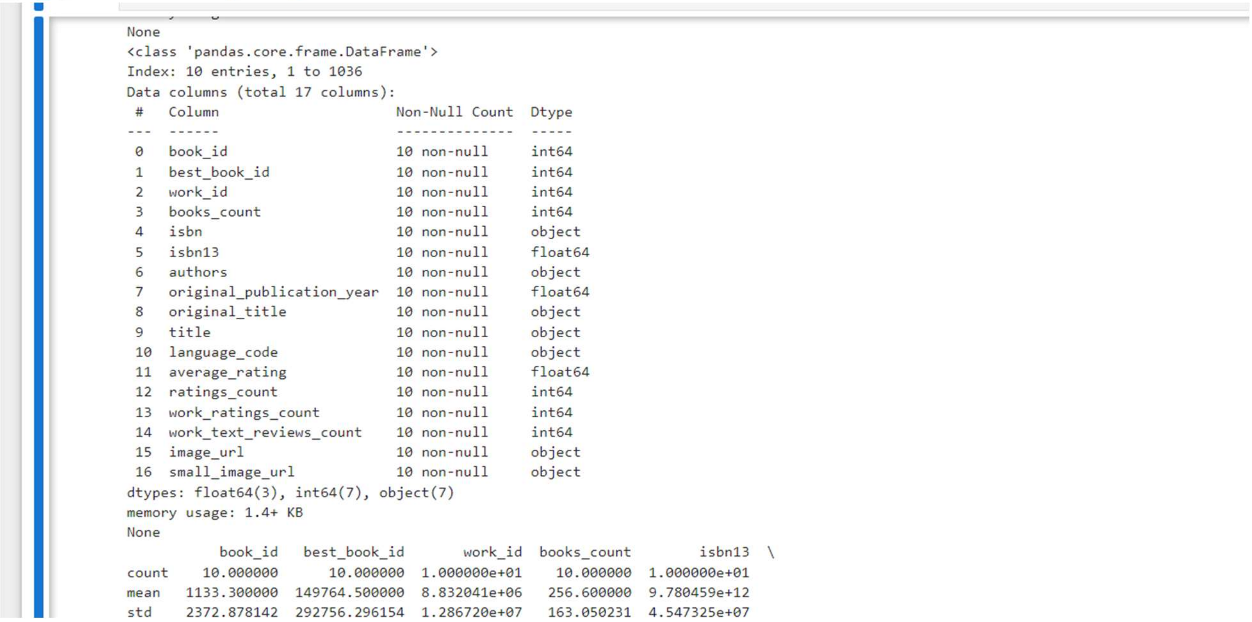
```
           book_id  goodreads_book_id  best_book_id       work_id  \
count  1354.000000       1.354000e+03  1.354000e+03  1.354000e+03
mean   4453.584195       5.951852e+06  6.120589e+06  8.707028e+06
std    2894.277455       6.664595e+06  6.935008e+06  9.813696e+06
min       1.000000       1.000000e+00  1.000000e+00  1.150000e+02
25%    1860.250000       1.537868e+05  1.537962e+05  1.375035e+06
50%    4177.500000       3.305318e+06  3.422646e+06  4.005716e+06
75%    6814.500000       9.917380e+06  1.019388e+07  1.435717e+07
max    9955.000000       3.207567e+07  3.360215e+07  4.963819e+07

       books_count        isbn13  original_publication_year  average_rating  \
count  1354.000000  1.310000e+03                1351.000000     1354.000000
mean     50.330871  9.766700e+12                2003.422650        3.999357
std      61.338867  3.572069e+11                  16.779301        0.224263
min       1.000000  7.678361e+10                1868.000000        3.230000
25%      22.000000  9.780152e+12                2003.000000        3.850000
50%      37.000000  9.780440e+12                2008.000000        4.000000
75%      58.000000  9.780805e+12                2011.000000        4.160000
max    1314.000000  9.788424e+12                2017.000000        4.740000

       ratings_count  work_ratings_count  work_text_reviews_count  \
count   1.354000e+03        1.354000e+03              1354.000000
mean    9.160429e+04        9.915569e+04              5151.093058
std     2.871266e+05        3.023637e+05             10730.335273
min     6.221000e+03        8.833000e+03                49.000000
25%     1.759325e+04        1.918150e+04              1162.500000
50%     2.943000e+04        3.255150e+04              2208.000000
75%     6.073800e+04        6.681275e+04              4690.750000
max     4.780653e+06        4.942365e+06            155254.000000
```



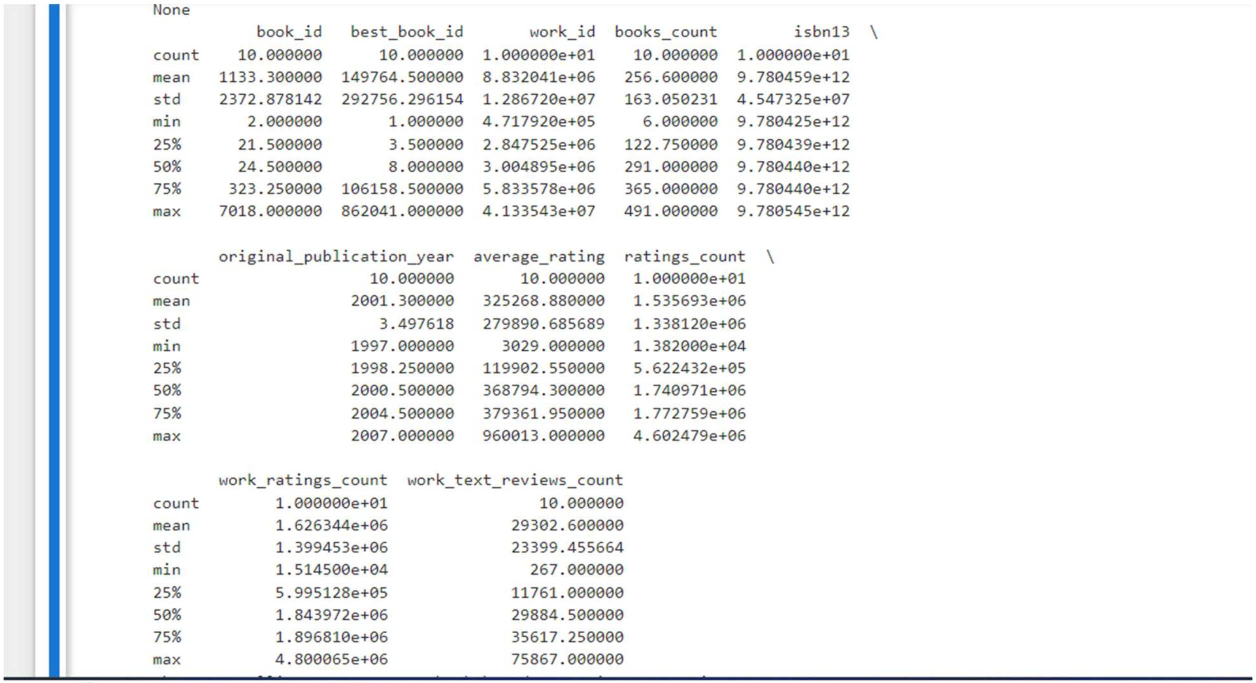
```
           ratings_1      ratings_2      ratings_3     ratings_4     ratings_5
count    1354.000000    1354.000000    1354.000000  1.354000e+03  1.354000e+03
mean     2297.409158    5005.615953   17528.918021  3.060591e+04  4.371784e+04
std     13708.507239   16259.838433   43549.306920  8.427851e+04  1.610638e+05
min        33.000000     133.000000     826.000000  1.660000e+03  2.005000e+03
25%       306.000000     978.000000    4140.500000  6.360500e+03  6.981500e+03
50%       619.000000    1732.500000    6557.000000  1.079550e+04  1.182650e+04
75%      1355.000000    3644.500000   13312.250000  2.227500e+04  2.612400e+04
max    456191.000000  436802.000000  793319.000000  1.481305e+06  3.011543e+06
Index(['book_id', 'goodreads_book_id', 'best_book_id', 'work_id',
       'books_count', 'isbn', 'isbn13', 'authors', 'original_publication_year',
       'original_title', 'title', 'language_code', 'average_rating',
       'ratings_count', 'work_ratings_count', 'work_text_reviews_count',
       'ratings_1', 'ratings_2', 'ratings_3', 'ratings_4', 'ratings_5',
       'image_url', 'small_image_url'],
      dtype='object')
```



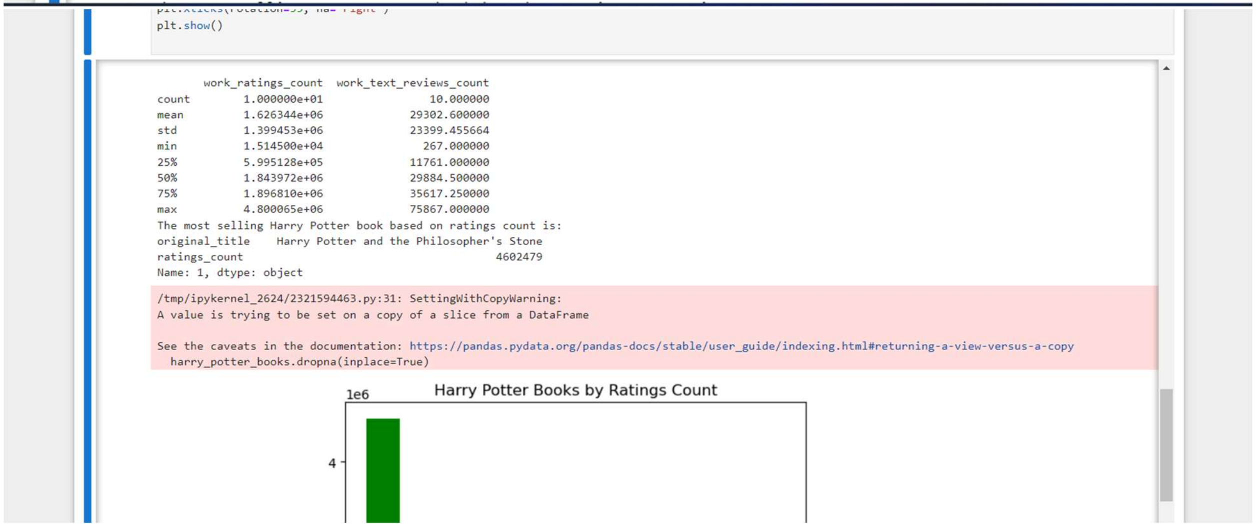
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1354 entries, 0 to 1353
Data columns (total 23 columns):
 #   Column                     Non-Null Count  Dtype  
---  ------                     --------------  -----  
 0   book_id                    1354 non-null   int64  
 1   goodreads_book_id          1354 non-null   int64  
 2   best_book_id               1354 non-null   int64  
 3   work_id                    1354 non-null   int64  
 4   books_count                1354 non-null   int64  
 5   isbn                       1302 non-null   object 
 6   isbn13                     1310 non-null   float64
 7   authors                    1354 non-null   object 
 8   original_publication_year  1351 non-null   float64
 9   original_title             1302 non-null   object 
 10  title                      1354 non-null   object 
 11  language_code              1245 non-null   object 
 12  average_rating             1354 non-null   float64
 13  ratings_count              1354 non-null   int64  
 14  work_ratings_count         1354 non-null   int64  
 15  work_text_reviews_count    1354 non-null   int64  
 16  ratings_1                  1354 non-null   int64  
 17  ratings_2                  1354 non-null   int64  
 18  ratings_3                  1354 non-null   int64  
 19  ratings_4                  1354 non-null   int64  
 20  ratings_5                  1354 non-null   int64  
 21  image_url                  1354 non-null   object 
 22  small_image_url            1354 non-null   object 
dtypes: float64(3), int64(13), object(7)
memory usage: 243.4+ KB
```



```
None
<class 'pandas.core.frame.DataFrame'>
Index: 10 entries, 1 to 1036
Data columns (total 17 columns):
 #   Column                     Non-Null Count  Dtype  
---  ------                     --------------  -----  
 0   book_id                    10 non-null     int64  
 1   best_book_id               10 non-null     int64  
 2   work_id                    10 non-null     int64  
 3   books_count                10 non-null     int64  
 4   isbn                       10 non-null     object 
 5   isbn13                     10 non-null     float64
 6   authors                    10 non-null     object 
 7   original_publication_year  10 non-null     float64
 8   original_title             10 non-null     object 
 9   title                      10 non-null     object 
 10  language_code              10 non-null     object 
 11  average_rating             10 non-null     float64
 12  ratings_count              10 non-null     int64  
 13  work_ratings_count         10 non-null     int64  
 14  work_text_reviews_count    10 non-null     int64  
 15  image_url                  10 non-null     object 
 16  small_image_url            10 non-null     object 
dtypes: float64(3), int64(7), object(7)
memory usage: 1.4+ KB
None
           book_id   best_book_id       work_id  books_count        isbn13  \
count    10.000000      10.000000  1.000000e+01    10.000000  1.000000e+01   
mean   1133.300000  149764.500000  8.832041e+06   256.600000  9.780459e+12   
std    2372.878142  292756.296154  1.286720e+07   163.050231  4.547325e+07   
```



```
None
          book_id  best_book_id       work_id  books_count        isbn13  \
count    10.000000     10.000000  1.000000e+01    10.000000  1.000000e+01
mean   1133.300000 149764.500000  8.832041e+06   256.600000  9.780459e+12
std    2372.878142 292756.296154  1.286720e+07   163.050231  4.547325e+07
min       2.000000      1.000000  4.717920e+05     6.000000  9.780425e+12
25%      21.500000      3.500000  2.847525e+06   122.750000  9.780439e+12
50%      24.500000      8.000000  3.004895e+06   291.000000  9.780440e+12
75%     323.250000 106158.500000  5.833578e+06   365.000000  9.780440e+12
max    7018.000000 862041.000000  4.133543e+07   491.000000  9.780545e+12

       original_publication_year  average_rating  ratings_count  \
count                  10.000000       10.000000   1.000000e+01
mean                 2001.300000   325268.880000   1.535693e+06
std                     3.497618   279890.685689   1.338120e+06
min                  1997.000000     3029.000000   1.382000e+04
25%                  1998.250000   119902.550000   5.622432e+05
50%                  2000.500000   368794.300000   1.740971e+06
75%                  2004.500000   379361.950000   1.772759e+06
max                  2007.000000   960013.000000   4.602479e+06

       work_ratings_count  work_text_reviews_count
count        1.000000e+01                10.000000
mean         1.626344e+06             29302.600000
std          1.399453e+06             23399.455664
min          1.514500e+04               267.000000
25%          5.995128e+05             11761.000000
50%          1.843972e+06             29884.500000
75%          1.896810e+06             35617.250000
max          4.800065e+06             75867.000000
```

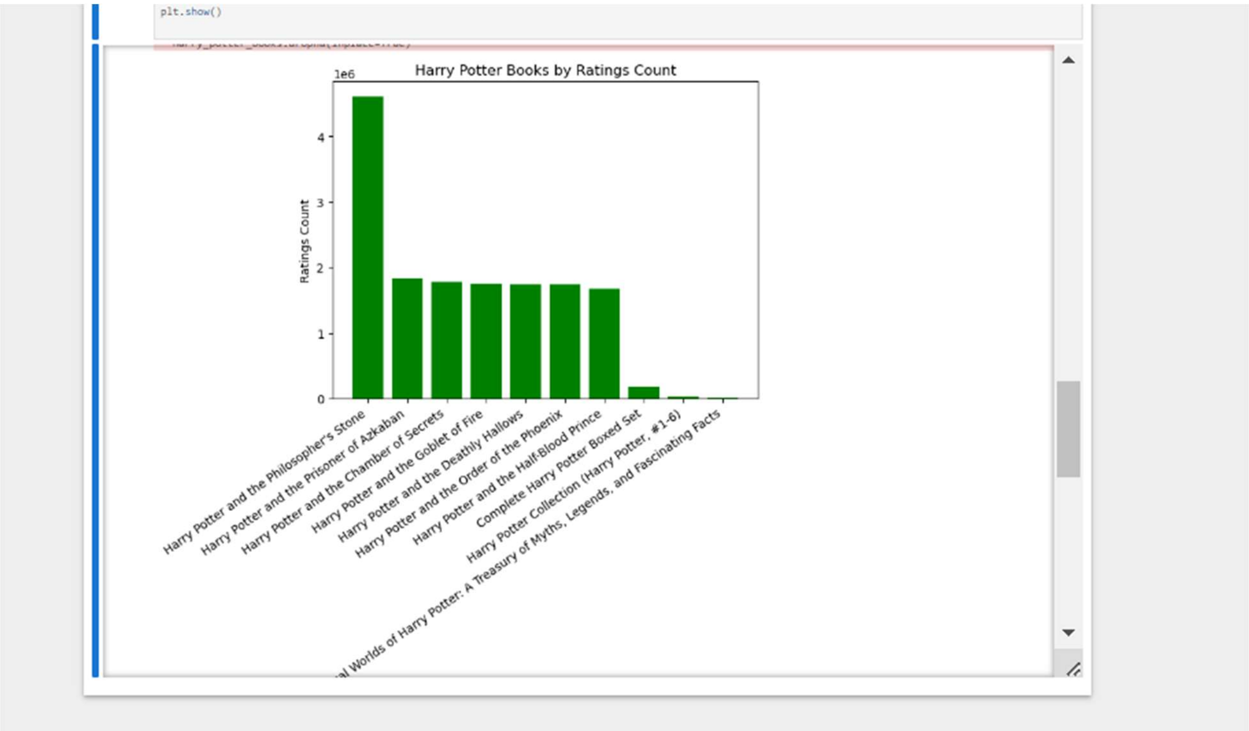


```
plt.xticks(rotation=35, ha='right')
plt.show()
```

```
       work_ratings_count  work_text_reviews_count
count        1.000000e+01                10.000000
mean         1.626344e+06             29302.600000
std          1.399453e+06             23399.455664
min          1.514500e+04               267.000000
25%          5.995128e+05             11761.000000
50%          1.843972e+06             29884.500000
75%          1.896810e+06             35617.250000
max          4.800065e+06             75867.000000
```

```
The most selling Harry Potter book based on ratings count is:
original_title    Harry Potter and the Philosopher's Stone
ratings_count                                      4602479
Name: 1, dtype: object
```

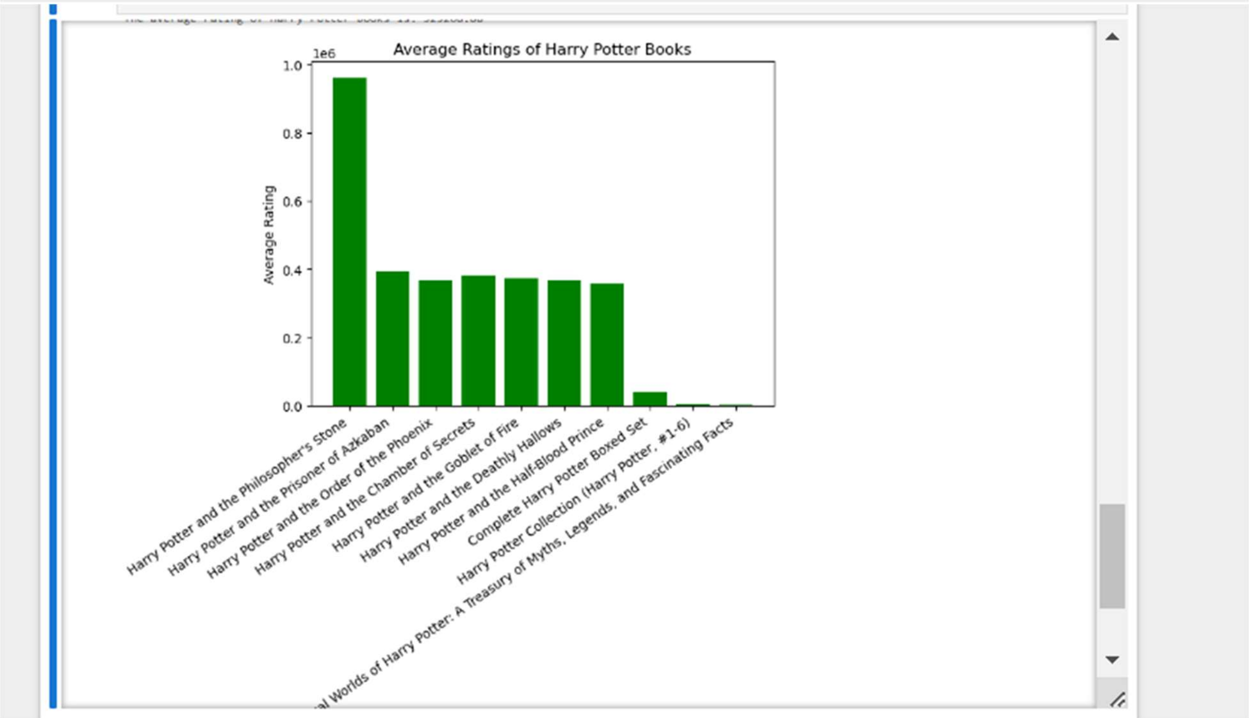
```
/tmp/ipykernel_2624/2321594463.py:31: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  harry_potter_books.dropna(inplace=True)
```



```
plt.show()
```

```
harry_potter_books.dropna(inplace=True)
```

Harry Potter Books by Ratings Count
1e6
Ratings Count
0
1
2
3
4
Harry Potter and the Philosopher's Stone
Harry Potter and the Prisoner of Azkaban
Harry Potter and the Chamber of Secrets
Harry Potter and the Goblet of Fire
Harry Potter and the Deathly Hallows
Harry Potter and the Order of the Phoenix
Harry Potter and the Half-Blood Prince
Complete Harry Potter Boxed Set
Harry Potter Collection (Harry Potter, #1-6)
al Worlds of Harry Potter: A Treasury of Myths, Legends, and Fascinating Facts



```
The average rating of Harry Potter books is: [illegible]
```

Average Ratings of Harry Potter Books
1e6
Average Rating
0.0
0.2
0.4
0.6
0.8
1.0
Harry Potter and the Philosopher's Stone
Harry Potter and the Prisoner of Azkaban
Harry Potter and the Order of the Phoenix
Harry Potter and the Chamber of Secrets
Harry Potter and the Goblet of Fire
Harry Potter and the Deathly Hallows
Harry Potter and the Half-Blood Prince
Complete Harry Potter Boxed Set
Harry Potter Collection (Harry Potter, #1-6)
al Worlds of Harry Potter: A Treasury of Myths, Legends, and Fascinating Facts