Home | Upload files · Mazenyasser223/C | how to upload project on github | Home | Untitled

localhost:8888/notebooks/Untitled.ipynb

jupyter Untitled Last Checkpoint: 12 hours ago

File Edit View Run Kernel Settings Help

Not Trusted

Code | Open in... | Python 3 (ipykernel)

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
df = pd.read_csv("books.csv")
df.head()
```

[2]:

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings_count	wo
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	...	4780653	
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	4602479	
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	...	3866839	
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	...	2346404	
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	...	1903563	

5 rows × 23 columns

[3]:
```
df.describe()
```

[3]:

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn13	original_publication_year	average_rating	ratings_count	work_ratings_c

23°C غائم غالبًا

ENG 9:27 PM 24-Apr-24

jupyter Untitled Last Checkpoint: 12 hours ago

File Edit View Run Kernel Settings Help

Not Trusted

Code

Open in... Python 3 (ipykernel)

[3]:
```
df.describe()
```

[3]:

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn13	original_publication_year	average_rating	ratings_count	work_ratings_c
count	1354.000000	1.354000e+03	1.354000e+03	1.354000e+03	1354.000000	1.310000e+03	1351.000000	1354.000000	1.354000e+03	1.354000
mean	4453.584195	5.951852e+06	6.120589e+06	8.707028e+06	50.330871	9.766700e+12	2003.422650	3.999357	9.160429e+04	9.915569
std	2894.277455	6.664595e+06	6.935008e+06	9.813696e+06	61.338867	3.572069e+11	16.779301	0.224263	2.871266e+05	3.023637
min	1.000000	1.000000e+00	1.000000e+00	1.150000e+02	1.000000	7.678361e+10	1868.000000	3.230000	6.221000e+03	8.833000
25%	1860.250000	1.537868e+05	1.537962e+05	1.375035e+06	22.000000	9.780152e+12	2003.000000	3.850000	1.759325e+04	1.918150
50%	4177.500000	3.305318e+06	3.422646e+06	4.005716e+06	37.000000	9.780440e+12	2008.000000	4.000000	2.943000e+04	3.255150
75%	6814.500000	9.917380e+06	1.019388e+07	1.435717e+07	58.000000	9.780805e+12	2011.000000	4.160000	6.073800e+04	6.681275
max	9955.000000	3.207567e+07	3.360215e+07	4.963819e+07	1314.000000	9.788424e+12	2017.000000	4.740000	4.780653e+06	4.942365

[4]:
```
df.info()
```

```
<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
```

jupyter Untitled Last Checkpoint: 12 hours ago

File Edit View Run Kernel Settings Help

Not Trusted

Code

Open in... Python 3 (ipykernel)

```
 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
```

[5]:
```
df.isnull().sum()
```

[5]:
```
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
```

[6]:
```
df.dropna(inplace= True)
```

[7]:
```
df.duplicated().sum()
```

jupyter Untitled Last Checkpoint: 12 hours ago

File Edit View Run Kernel Settings Help

Not Trusted

Code

Open in... Python 3 (ipykernel)

```
image_url                0
small_image_url          0
dtype: int64
```

[6]:
```
df.dropna(inplace= True)
```

[7]:
```
df.duplicated().sum()
```

[7]:
```
0
```

[8]:
```
cols = ['title', 'authors', 'original_publication_year', 'average_rating', 'ratings_count']
df = df[cols]


authors = df.authors.unique()
jkr = [i for i in authors if i.find('J.K. Rowling') != -1]
jkr_books = df.loc[df.authors.isin(jkr)]

all_titles = jkr_books.title.unique()
hp_titles = [i for i in all_titles if i.find('Harry Potter and ') != -1]

hp_books = jkr_books.loc[jkr_books.title.isin(hp_titles)]
hp_books = hp_books.sort_values(by=['original_publication_year'], ascending=True)

hp_books
```

[8]:

	title	authors	original_publication_year	average_rating	ratings_count
1	Harry Potter and the Sorcerer's Stone (Harry P...	J.K. Rowling, Mary GrandPré	1997.0	4.44	4602479
9	Harry Potter and the Chamber of Secrets (Harry...	J.K. Rowling, Mary GrandPré	1998.0	4.37	1779331
6	Harry Potter and the Prisoner of Azkaban (Harr...	J.K. Rowling, Mary GrandPré, Rufus Beck	1999.0	4.53	1832823
10	Harry Potter and the Goblet of Fire (Harry Pot...	J.K. Rowling, Mary GrandPré	2000.0	4.53	1753043
8	Harry Potter and the Order of the Phoenix (Har...	J.K. Rowling, Mary GrandPré	2003.0	4.46	1735368

jupyter Untitled Last Checkpoint: 12 hours ago

File Edit View Run Kernel Settings Help

Not Trusted

Code

Open in... Python 3 (ipykernel)

```
authors = df.authors.unique()
jkr = [i for i in authors if i.find('J.K. Rowling') != -1]
jkr_books = df.loc[df.authors.isin(jkr)]

all_titles = jkr_books.title.unique()
hp_titles = [i for i in all_titles if i.find('Harry Potter and ') != -1]

hp_books = jkr_books.loc[jkr_books.title.isin(hp_titles)]
hp_books = hp_books.sort_values(by=['original_publication_year'], ascending=True)

hp_books
```

[8]:

	title	authors	original_publication_year	average_rating	ratings_count
1	Harry Potter and the Sorcerer's Stone (Harry P...	J.K. Rowling, Mary GrandPré	1997.0	4.44	4602479
9	Harry Potter and the Chamber of Secrets (Harry...	J.K. Rowling, Mary GrandPré	1998.0	4.37	1779331
6	Harry Potter and the Prisoner of Azkaban (Harr...	J.K. Rowling, Mary GrandPré, Rufus Beck	1999.0	4.53	1832823
10	Harry Potter and the Goblet of Fire (Harry Pot...	J.K. Rowling, Mary GrandPré	2000.0	4.53	1753043
8	Harry Potter and the Order of the Phoenix (Har...	J.K. Rowling, Mary GrandPré	2003.0	4.46	1735368
12	Harry Potter and the Half-Blood Prince (Harry ...	J.K. Rowling, Mary GrandPré	2005.0	4.54	1678823
11	Harry Potter and the Deathly Hallows (Harry Po...	J.K. Rowling, Mary GrandPré	2007.0	4.61	1746574

[9]:

```
weighted_avg = (hp_books.average_rating * hp_books.ratings_count).sum() / hp_books.ratings_count.sum()
titles = hp_books.title.tolist()

sns.set_style("whitegrid")
plt.figure(figsize=(12, 6))
sns.barplot(x=titles, y=hp_books.average_rating, palette="viridis")
plt.xlabel('Title', fontsize=12)
plt.ylabel('Average rating', fontsize=12)
plt.ylim((4.3, 4.7))
```

Home | Upload files · Mazenyasser223/C | how to upload project on github | Home | Untitled

localhost:8888/notebooks/Untitled.ipynb

jupyter Untitled Last Checkpoint: 12 hours ago

File Edit View Run Kernel Settings Help

Not Trusted

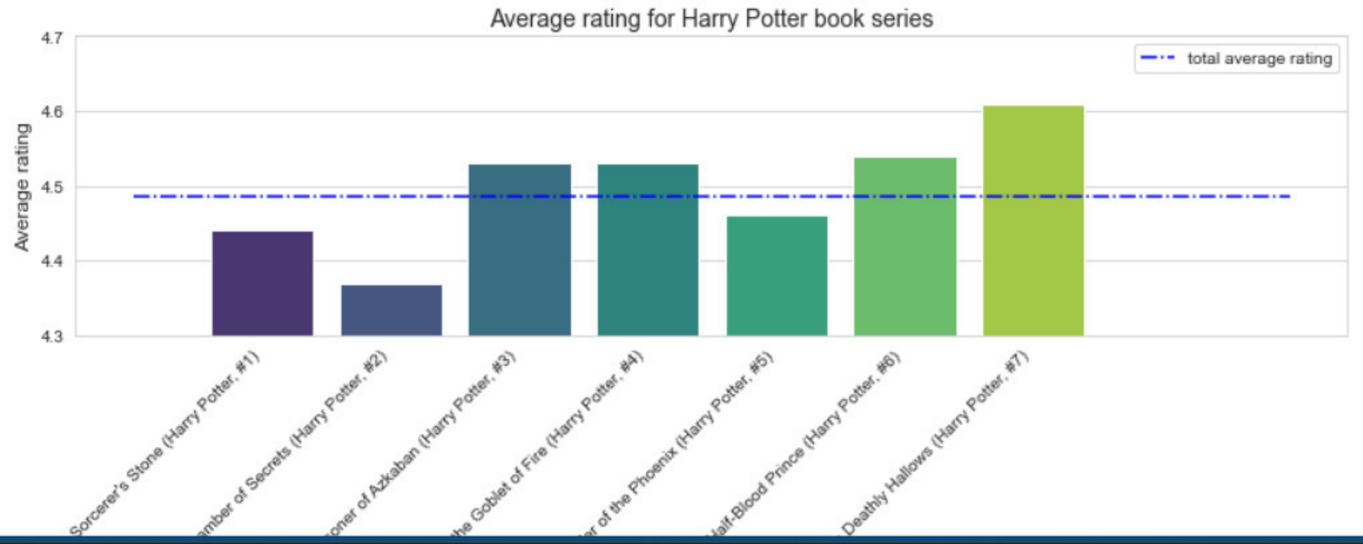
Code | Open in... | Python 3 (ipykernel)

[9]:
```
weighted_avg = (hp_books.average_rating * hp_books.ratings_count).sum() / hp_books.ratings_count.sum()
titles = hp_books.title.tolist()

sns.set_style("whitegrid")
plt.figure(figsize=(12, 6))
sns.barplot(x=titles, y=hp_books.average_rating, palette="viridis")
plt.xlabel('Title', fontsize=12)
plt.ylabel('Average rating', fontsize=12)
plt.ylim((4.3, 4.7))

plt.hlines(weighted_avg, xmin=-1, xmax=8, color='blue', linestyles='dashdot', label='total average rating')
plt.title('Average rating for Harry Potter book series', fontsize=14)
plt.xticks(rotation=45, ha='right')

plt.tight_layout()
plt.legend()
plt.show()
```



23°C غائم غالبًا | ENG | 9:28 PM 24-Apr-24

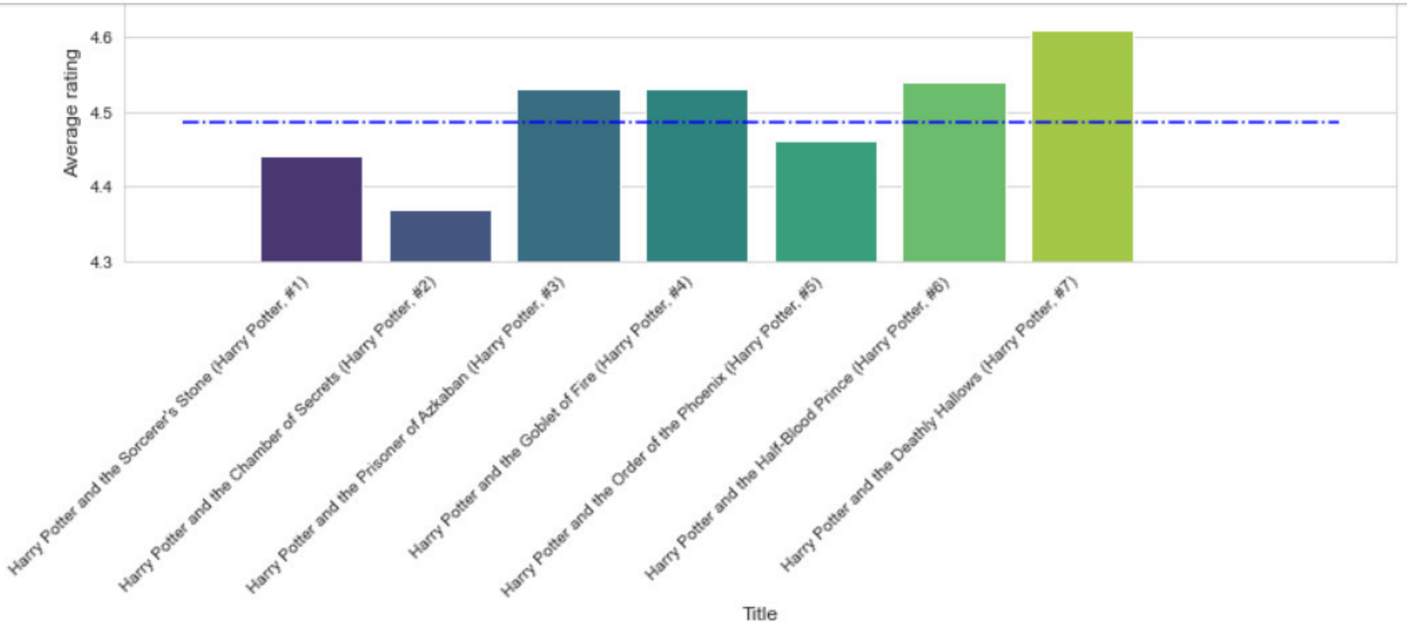
jupyter Untitled Last Checkpoint: 12 hours ago

File Edit View Run Kernel Settings Help

Not Trusted

Code

Open in... Python 3 (ipykernel)



[10]:
```
print("Average Ratings : ",weighted_avg)
```

```
Average Ratings :  4.486117244334695
```

[ ]: