

netflix-1

November 20, 2023

```
[3]: import pandas as pd
from tqdm.auto import tqdm

def read_all_sheets_from_excel(path: str) -> dict:
    all_sheets = pd.read_excel(path, sheet_name=None)
    for name, df in tqdm(all_sheets.items(), desc='Loading sheets'):
        df = df.rename(columns=lambda x: x.strip() if isinstance(x, str) else x)
        df = df.apply(lambda col: col.str.strip() if col.dtype == "object" else
↳col, axis=0)
        all_sheets[name] = df
    return all_sheets

dataframes = read_all_sheets_from_excel('C:/Users/futak/.ipynb_checkpoints/
↳Netflix Dataset Latest 2021.xlsx')
print('Dataframes in this file:', ', '.join(dataframes.keys()))

for k, v in dataframes.items():
    print(f'\n{k} - First 5 rows:')
    print(v.head())
```

Loading sheets: 0%| | 0/1 [00:00<?, ?it/s]

Dataframes in this file: FlixGem.com dataset

FlixGem.com dataset - First 5 rows:

	Title	Genre
0	Lets Fight Ghost	Crime, Drama, Fantasy, Horror, Romance \
1	HOW TO BUILD A GIRL	Comedy
2	The Con-Heartist	Comedy, Romance
3	Gleboka woda	Drama
4	Only a Mother	Drama

	Tags	Languages
0	Comedy Programmes,Romantic TV Comedies,Horror ...	Swedish, Spanish \
1	Dramas,Comedies,Films Based on Books,British	English
2	Romantic Comedies,Comedies,Romantic Films,Thai...	Thai
3	TV Dramas,Polish TV Shows,Social Issue TV Dramas	Polish
4	Social Issue Dramas,Dramas,Movies Based on Boo...	Swedish

	Series or Movie	Hidden Gem Score	
0	Series	4.3	\
1	Movie	7.0	
2	Movie	8.6	
3	Series	8.7	
4	Movie	8.3	

		Country Availability	Runtime	
0		Thailand	< 30 minutes	\
1		Canada	1-2 hour	
2		Thailand	> 2 hrs	
3		Poland	< 30 minutes	
4	Lithuania, Poland, France, Italy, Spain, Greece, Bel...		1-2 hour	

	Director	Writer	...
0	Tomas Alfredson	John Ajvide Lindqvist	\
1	Coky Giedroyc	Caitlin Moran	...
2	Mez Tharatorn	Pattaranad Bhiboonsawade, Mez Tharatorn, Thods...	...
3	NaN	NaN	...
4	Alf Sjöberg	Ivar Lo-Johansson	...

	Netflix Release Date	Production House	
0	2021-03-04	Canal+, Sandrew Metronome	\
1	2021-03-04	Film 4, Monumental Pictures, Lionsgate	
2	2021-03-03	NaN	
3	2021-03-03	NaN	
4	2021-03-03	NaN	

	Netflix Link	
0	https://www.netflix.com/watch/81415947	\
1	https://www.netflix.com/watch/81041267	
2	https://www.netflix.com/watch/81306155	
3	https://www.netflix.com/watch/81307527	
4	https://www.netflix.com/watch/81382068	

	IMDb Link	
0	https://www.imdb.com/title/tt1139797	\
1	https://www.imdb.com/title/tt4193072	
2	https://www.imdb.com/title/tt13393728	
3	https://www.imdb.com/title/tt2300049	
4	https://www.imdb.com/title/tt0041155	

	Summary	IMDb Votes	
0	A med student with a supernatural gift tries t...	205926.0	\
1	When nerdy Johanna moves to London, things get...	2838.0	
2	After her ex-boyfriend cons her out of a large...	131.0	
3	A group of social welfare workers led by their...	47.0	

4 An unhappily married farm worker struggling to... 88.0

Image

```
0 https://occ-0-4708-64.1.nflxso.net/dnm/api/v6/... \
1 https://occ-0-1081-999.1.nflxso.net/dnm/api/v6...
2 https://occ-0-2188-64.1.nflxso.net/dnm/api/v6/...
3 https://occ-0-2508-2706.1.nflxso.net/dnm/api/v...
4 https://occ-0-2851-41.1.nflxso.net/dnm/api/v6/...
```

Poster

```
0 https://m.media-amazon.com/images/M/MV5B0WM4NT... \
1 https://m.media-amazon.com/images/M/MV5BZGUyN2...
2 https://m.media-amazon.com/images/M/MV5B0DAz0G...
3 https://m.media-amazon.com/images/M/MV5BMTc0Nz...
4 https://m.media-amazon.com/images/M/MV5BMjVmMz...
```

TMDb Trailer Trailer Site

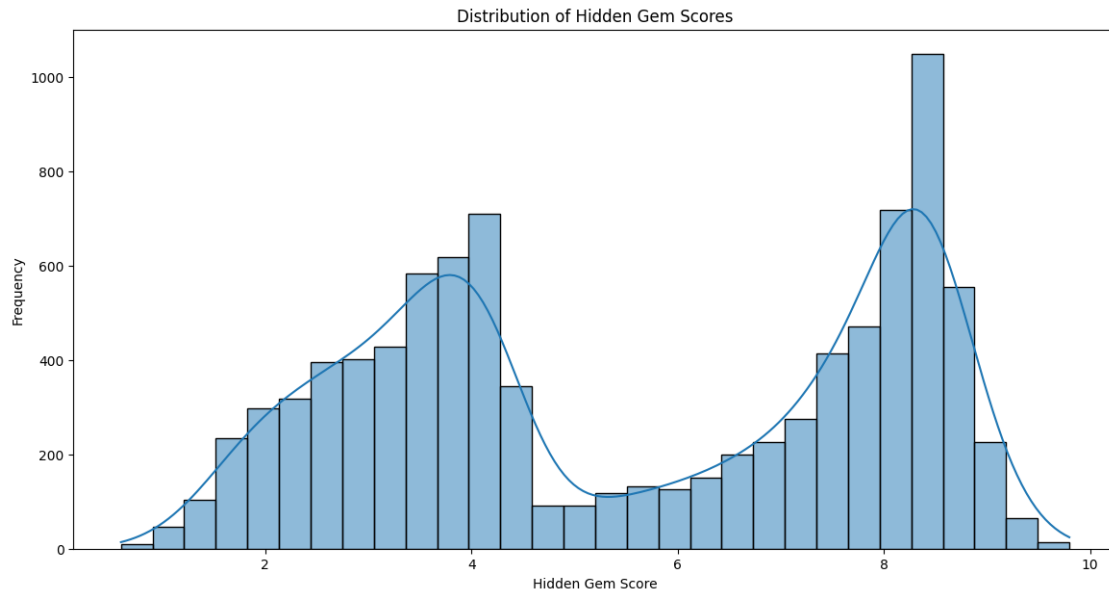
```
0 https://www.youtube.com/watch?v=LqB6XJix-dM      YouTube
1 https://www.youtube.com/watch?v=eIbcxPy4okQ      YouTube
2 https://www.youtube.com/watch?v=md3CmFLGK6Y      YouTube
3 https://www.youtube.com/watch?v=5kyF2vy63r0      YouTube
4 https://www.youtube.com/watch?v=H0itWKFwMpQ      YouTube
```

[5 rows x 29 columns]

```
[7]: import matplotlib.pyplot as plt
import seaborn as sns

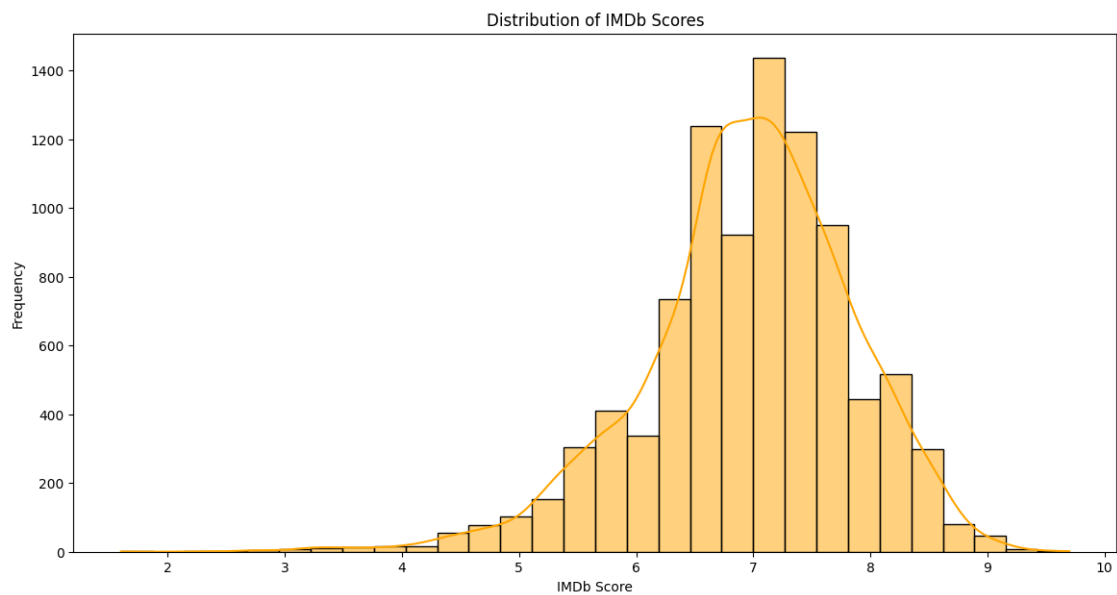
flixgem_df = dataframes['FlixGem.com dataset']

plt.figure(figsize=(14, 7))
sns.histplot(flixgem_df['Hidden Gem Score'].dropna(), bins=30, kde=True)
plt.title('Distribution of Hidden Gem Scores')
plt.xlabel('Hidden Gem Score')
plt.ylabel('Frequency')
plt.show()
```



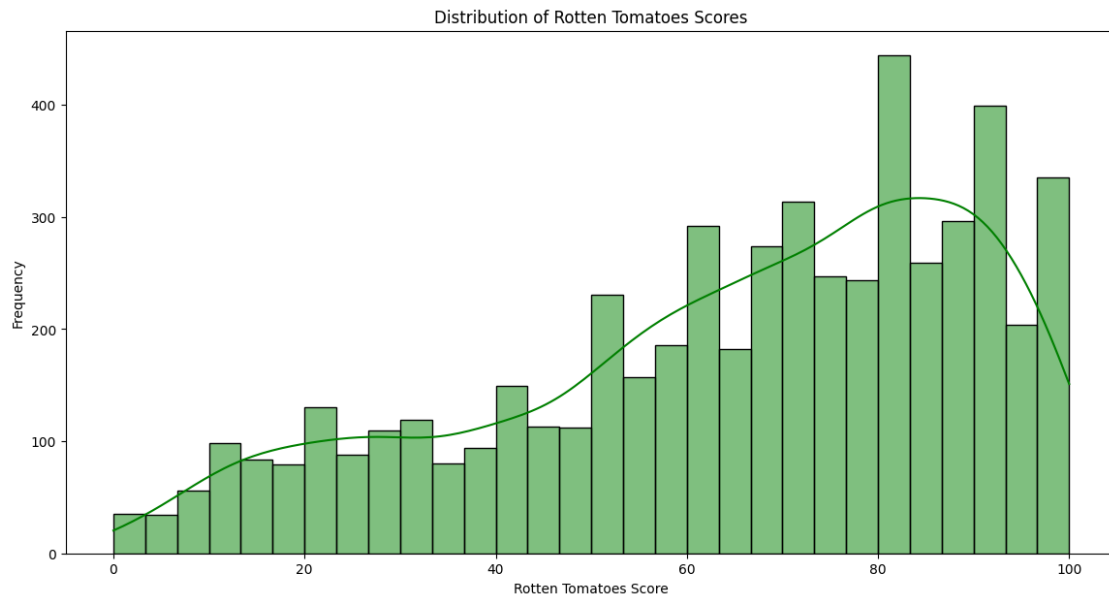
```
[10]: plt.figure(figsize=(14, 7))

sns.histplot(flixgem_df['IMDb Score'].dropna(), bins=30, kde=True,
             color='orange')
plt.title('Distribution of IMDb Scores')
plt.xlabel('IMDb Score')
plt.ylabel('Frequency')
plt.show()
```



```
[9]: plt.figure(figsize=(14, 7))

sns.histplot(flixgem_df['Rotten Tomatoes Score'].dropna(), bins=30, kde=True,
             color='green')
plt.title('Distribution of Rotten Tomatoes Scores')
plt.xlabel('Rotten Tomatoes Score')
plt.ylabel('Frequency')
plt.show()
```



```
[11]: plt.figure(figsize=(14, 7))

sns.histplot(flixgem_df['Metacritic Score'].dropna(), bins=30, kde=True,
             color='purple')
plt.title('Distribution of Metacritic Scores')
plt.xlabel('Metacritic Score')
plt.ylabel('Frequency')
plt.show()
```



```
[12]: flixgem_df.describe()
```

```
[12]:
```

	Hidden Gem Score	IMDb Score	Rotten Tomatoes Score	Metacritic Score	
count	9415.000000	9417.000000	5445.000000	4082.000000	\
mean	5.540733	6.955517	64.691276	58.113425	
min	0.600000	1.600000	0.000000	6.000000	
25%	3.400000	6.500000	49.000000	46.000000	
50%	5.300000	7.000000	70.000000	59.000000	
75%	8.100000	7.500000	85.000000	71.000000	
max	9.800000	9.700000	100.000000	100.000000	
std	2.447462	0.899681	25.269466	17.143187	

	Awards Received	Awards Nominated For	Boxoffice	
count	5226.000000	6376.000000	3.754000e+03	\
mean	9.735936	16.035602	4.845788e+07	
min	1.000000	1.000000	7.200000e+01	
25%	1.250000	2.000000	1.243910e+06	
50%	4.000000	6.000000	2.250466e+07	
75%	9.000000	15.000000	6.425437e+07	
max	300.000000	386.000000	6.593639e+08	
std	19.524116	32.209094	7.243625e+07	

	Release Date	Netflix Release Date	
count	9217	9425	\
mean	2010-09-14 00:48:44.682651648	2017-11-19 08:53:13.209549056	
min	1913-10-27 00:00:00	2015-04-14 00:00:00	
25%	2007-08-03 00:00:00	2015-08-13 00:00:00	

50%	2014-10-10 00:00:00	2017-12-29 00:00:00
75%	2018-01-03 00:00:00	2019-10-05 00:00:00
max	2021-11-12 00:00:00	2021-03-04 00:00:00
std	NaN	NaN

	IMDb Votes
count	9.415000e+03
mean	6.014725e+04
min	5.000000e+00
25%	9.695000e+02
50%	6.602000e+03
75%	5.098700e+04
max	2.354197e+06
std	1.463837e+05

```
[13]: flixgem_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9425 entries, 0 to 9424
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Title                                9399 non-null   object
1   Genre                                9400 non-null   object
2   Tags                                 9389 non-null   object
3   Languages                            9255 non-null   object
4   Series or Movie                      9425 non-null   object
5   Hidden Gem Score                     9415 non-null   float64
6   Country Availability                 9414 non-null   object
7   Runtime                             9424 non-null   object
8   Director                             7120 non-null   object
9   Writer                               7615 non-null   object
10  Actors                               9314 non-null   object
11  View Rating                          6827 non-null   object
12  IMDb Score                           9417 non-null   float64
13  Rotten Tomatoes Score                5445 non-null   float64
14  Metacritic Score                     4082 non-null   float64
15  Awards Received                      5226 non-null   float64
16  Awards Nominated For                 6376 non-null   float64
17  Boxoffice                            3754 non-null   float64
18  Release Date                         9217 non-null   datetime64[ns]
19  Netflix Release Date                 9425 non-null   datetime64[ns]
20  Production House                     4393 non-null   object
21  Netflix Link                         9425 non-null   object
22  IMDb Link                            9101 non-null   object
23  Summary                              9420 non-null   object
24  IMDb Votes                           9415 non-null   float64
```

```

25 Image          9425 non-null object
26 Poster         8487 non-null object
27 TMDb Trailer   9421 non-null object
28 Trailer Site   9424 non-null object
dtypes: datetime64[ns](2), float64(8), object(19)
memory usage: 2.1+ MB

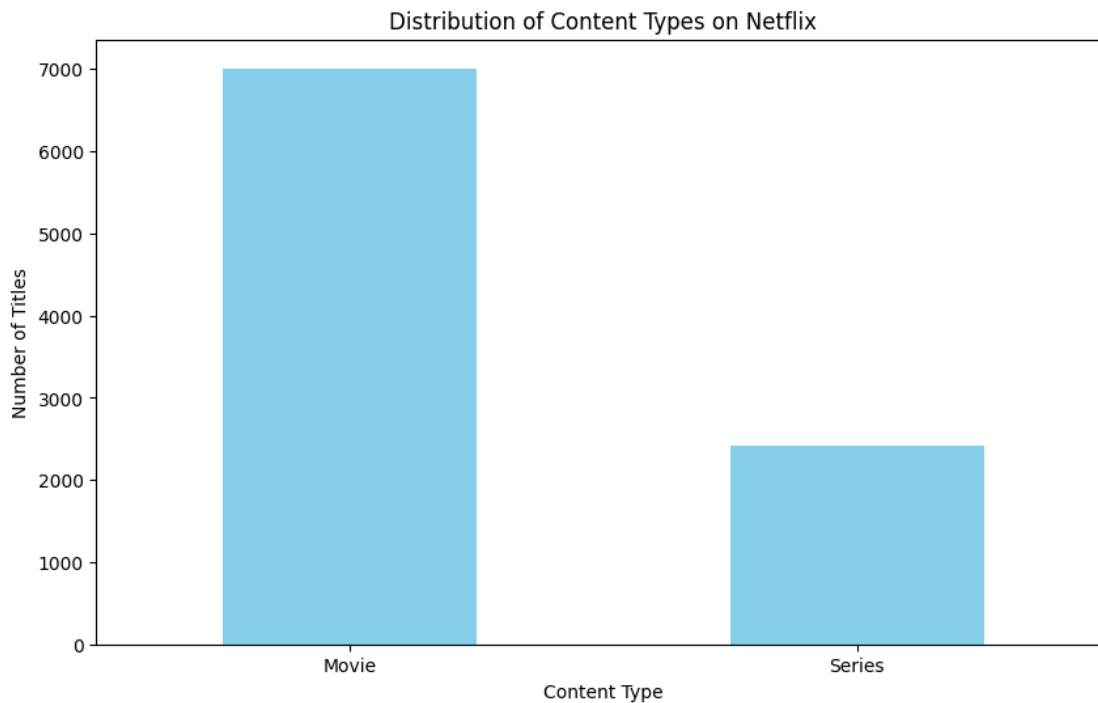
```

```

[14]: content_type_counts = flixgem_df['Series or Movie'].value_counts()

plt.figure(figsize=(10, 6))
content_type_counts.plot(kind='bar', color='skyblue')
plt.title('Distribution of Content Types on Netflix')
plt.xlabel('Content Type')
plt.ylabel('Number of Titles')
plt.xticks(rotation=0)
plt.show()

```



```

[15]: flixgem_df['Genre'].value_counts().head(10)

```

```

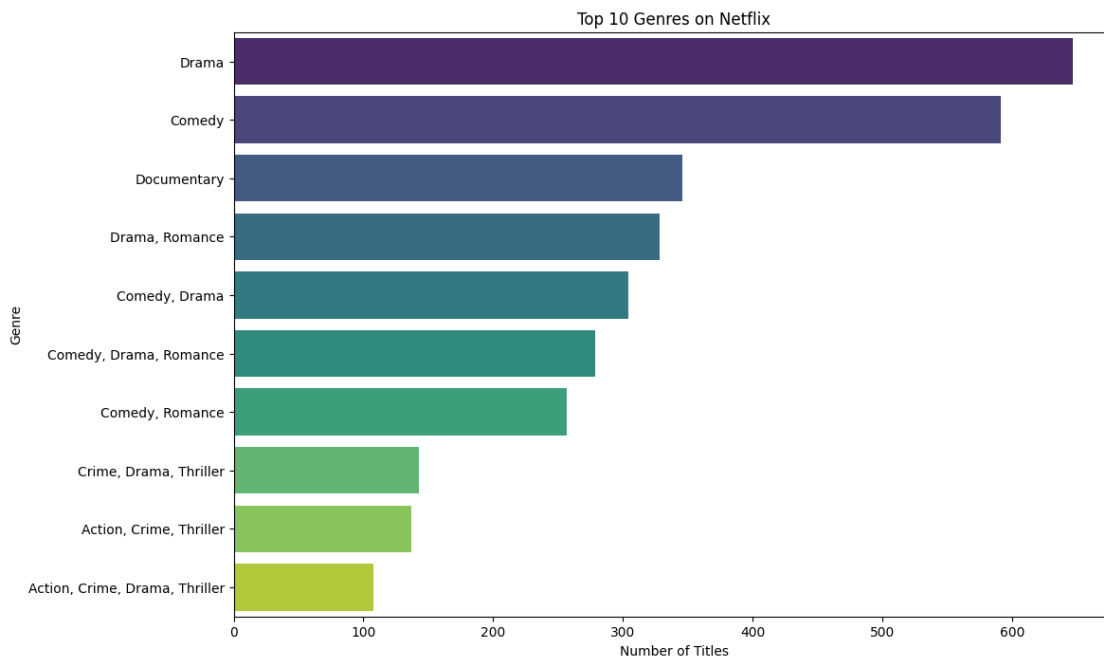
[15]: Genre
Drama          647
Comedy         591
Documentary    346
Drama, Romance 328

```


Comedy, Drama	304
Comedy, Drama, Romance	279
Comedy, Romance	257
Crime, Drama, Thriller	143
Action, Crime, Thriller	137
Action, Crime, Drama, Thriller	108

Name: count, dtype: int64

```
[16]: plt.figure(figsize=(12, 8))
sns.countplot(y='Genre', data=flixgem_df, order=flixgem_df['Genre'].
↳value_counts().index[:10], palette='viridis')
plt.title('Top 10 Genres on Netflix')
plt.xlabel('Number of Titles')
plt.ylabel('Genre')
plt.show()
```



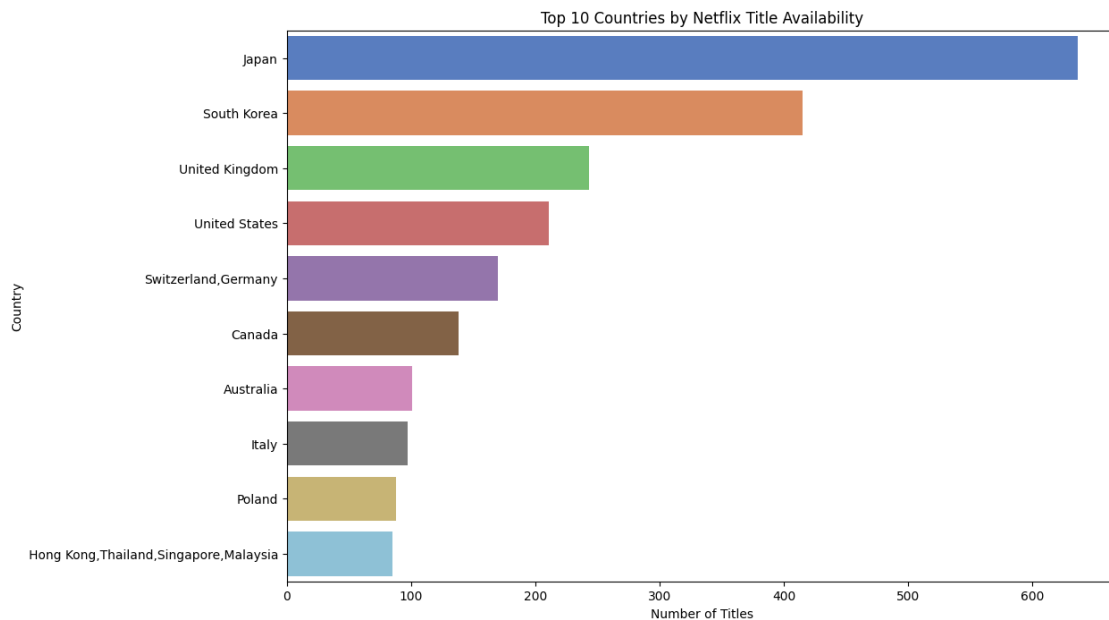
```
[17]: flixgem_df['Country Availability'].value_counts().head(10)
```

Country Availability	
Japan	637
South Korea	415
United Kingdom	243
United States	211
Switzerland, Germany	170
Canada	138

Australia	101
Italy	97
Poland	88
Hong Kong,Thailand,Singapore,Malaysia	85

Name: count, dtype: int64

```
[18]: plt.figure(figsize=(12, 8))
sns.countplot(y='Country Availability', data=flixgem_df,
              order=flixgem_df['Country Availability'].value_counts().index[:10],
              palette='muted')
plt.title('Top 10 Countries by Netflix Title Availability')
plt.xlabel('Number of Titles')
plt.ylabel('Country')
plt.show()
```



```
[19]: flixgem_df['Split Genres'] = flixgem_df['Genre'].str.split(', ')
exploded_genres = flixgem_df.explode('Split Genres')
average_imdb_scores = exploded_genres.groupby('Split Genres')['IMDb Score'].
    mean().sort_values(ascending=False)
average_imdb_scores.head(10)
```

```
[19]: Split Genres
Talk-Show      7.882353
News           7.557143
Documentary    7.416687
Reality-TV     7.360360
```

```

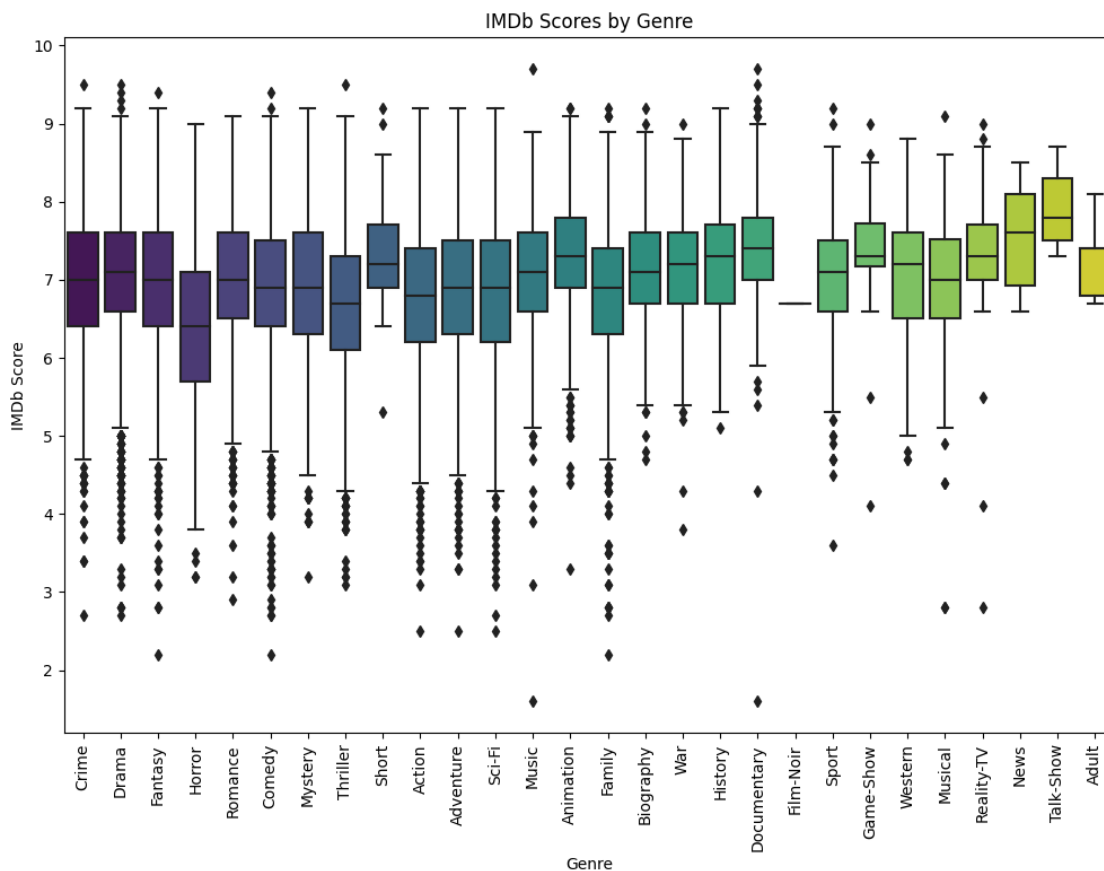
Game-Show      7.340625
Short          7.329193
Animation      7.320290
Adult          7.280000
History        7.225000
War            7.162411
Name: IMDb Score, dtype: float64

```

```

[21]: import seaborn as sns
plt.figure(figsize=(12, 8))
sns.boxplot(x='Split Genres', y='IMDb Score', data=exploded_genres,
            palette='viridis')
plt.title('IMDb Scores by Genre')
plt.xlabel('Genre')
plt.ylabel('IMDb Score')
plt.xticks(rotation=90)
plt.show()

```



```

[22]: exploded_genres.head()

```

[22]:

	Title	Genre
0	Lets Fight Ghost	Crime, Drama, Fantasy, Horror, Romance \
0	Lets Fight Ghost	Crime, Drama, Fantasy, Horror, Romance
0	Lets Fight Ghost	Crime, Drama, Fantasy, Horror, Romance
0	Lets Fight Ghost	Crime, Drama, Fantasy, Horror, Romance
0	Lets Fight Ghost	Crime, Drama, Fantasy, Horror, Romance

	Tags	Languages
0	Comedy Programmes,Romantic TV Comedies,Horror ...	Swedish, Spanish \
0	Comedy Programmes,Romantic TV Comedies,Horror ...	Swedish, Spanish
0	Comedy Programmes,Romantic TV Comedies,Horror ...	Swedish, Spanish
0	Comedy Programmes,Romantic TV Comedies,Horror ...	Swedish, Spanish
0	Comedy Programmes,Romantic TV Comedies,Horror ...	Swedish, Spanish

	Series or Movie	Hidden Gem Score	Country	Availability	Runtime
0	Series	4.3		Thailand	< 30 minutes \
0	Series	4.3		Thailand	< 30 minutes
0	Series	4.3		Thailand	< 30 minutes
0	Series	4.3		Thailand	< 30 minutes
0	Series	4.3		Thailand	< 30 minutes

	Director	Writer	...	Production House
0	Tomas Alfredson	John Ajvide Lindqvist	...	Canal+, Sandrew Metronome \
0	Tomas Alfredson	John Ajvide Lindqvist	...	Canal+, Sandrew Metronome
0	Tomas Alfredson	John Ajvide Lindqvist	...	Canal+, Sandrew Metronome
0	Tomas Alfredson	John Ajvide Lindqvist	...	Canal+, Sandrew Metronome
0	Tomas Alfredson	John Ajvide Lindqvist	...	Canal+, Sandrew Metronome

	Netflix Link
0	https://www.netflix.com/watch/81415947 \
0	https://www.netflix.com/watch/81415947
0	https://www.netflix.com/watch/81415947
0	https://www.netflix.com/watch/81415947
0	https://www.netflix.com/watch/81415947

	IMDb Link
0	https://www.imdb.com/title/tt1139797 \
0	https://www.imdb.com/title/tt1139797
0	https://www.imdb.com/title/tt1139797
0	https://www.imdb.com/title/tt1139797
0	https://www.imdb.com/title/tt1139797

	Summary	IMDb Votes
0	A med student with a supernatural gift tries t...	205926.0 \
0	A med student with a supernatural gift tries t...	205926.0
0	A med student with a supernatural gift tries t...	205926.0
0	A med student with a supernatural gift tries t...	205926.0

0 A med student with a supernatural gift tries t... 205926.0

Image

```
0 https://occ-0-4708-64.1.nflxso.net/dnm/api/v6/... \
0 https://occ-0-4708-64.1.nflxso.net/dnm/api/v6/...
0 https://occ-0-4708-64.1.nflxso.net/dnm/api/v6/...
0 https://occ-0-4708-64.1.nflxso.net/dnm/api/v6/...
0 https://occ-0-4708-64.1.nflxso.net/dnm/api/v6/...
```

Poster

```
0 https://m.media-amazon.com/images/M/MV5BOWM4NT... \
0 https://m.media-amazon.com/images/M/MV5BOWM4NT...
0 https://m.media-amazon.com/images/M/MV5BOWM4NT...
0 https://m.media-amazon.com/images/M/MV5BOWM4NT...
0 https://m.media-amazon.com/images/M/MV5BOWM4NT...
```

TMDb Trailer Trailer Site Split Genres

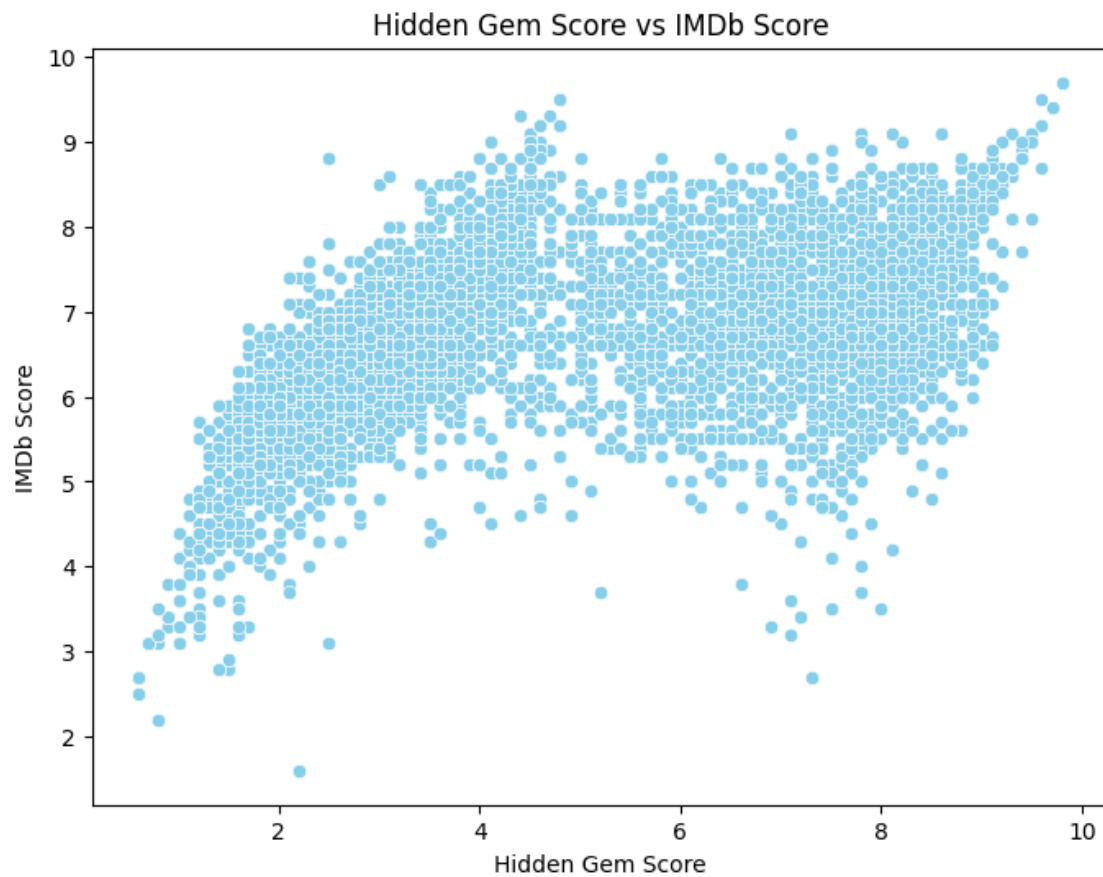
```
0 https://www.youtube.com/watch?v=LqB6XJix-dM YouTube Crime
0 https://www.youtube.com/watch?v=LqB6XJix-dM YouTube Drama
0 https://www.youtube.com/watch?v=LqB6XJix-dM YouTube Fantasy
0 https://www.youtube.com/watch?v=LqB6XJix-dM YouTube Horror
0 https://www.youtube.com/watch?v=LqB6XJix-dM YouTube Romance
```

[5 rows x 30 columns]

```
[23]: correlation = exploded_genres[['Hidden Gem Score', 'IMDb Score']].corr()
      print(correlation)
```

	Hidden Gem Score	IMDb Score
Hidden Gem Score	1.000000	0.402152
IMDb Score	0.402152	1.000000

```
[24]: plt.figure(figsize=(8, 6))
      sns.scatterplot(x='Hidden Gem Score', y='IMDb Score', data=exploded_genres,
                      color='skyblue')
      plt.title('Hidden Gem Score vs IMDb Score')
      plt.xlabel('Hidden Gem Score')
      plt.ylabel('IMDb Score')
      plt.show()
```



[]: