

NETFLIX_ANALYSIS

import libraries

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

In [2]: os.chdir('D:\\Hack_Challenge\\Project\\git\\SBSPS-Challenge-5374-OTT-Platform-Analysis-Tool\\DATASETS')

In [3]: netflix = pd.read_csv('netflix.csv')

In [4]: netflix.head()

Out[4]:
  show_id  type  title  director  cast  country  date_added  release_year  rating  duration  listed_in  description
0      s1    TV Show    3%      NaN  João Miguel, Bianca Comparato, Michel Gomes, R...  Brazil  August 14, 2020  2020  TV-MA  4 Seasons  International TV Shows, TV Dramas, TV Sci-Fi &...  In a future where the elite inhabit an island ...
1      s2    Movie    7.19  Jorge Michel Grau  Demian Bichir, Héctor Bonilla, Oscar Semán, ...  Mexico  December 23, 2016  2016  TV-MA  93 min  Dramas, International Movies  After a devastating earthquake his Mexico Cit...
2      s3    Movie    23.59  Gilbert Chan  Tedd Chan, Stella Chung, Henley Hii, Lawrence ...  Singapore  December 20, 2011  2011  R  78 min  Horror Movies, International Movies  When an army recruit is found dead, his fellow...
3      s4    TV Show    9  Shane Acker  Elijah Wood, John C. Reilly, Jennifer Connelly...  United States  November 16, 2017  2009  PG-13  80 min  Action & Adventure, Independent Movies, Sci-Fi...  In a postapocalyptic world, rag-doll robots hi...
4      s5    Movie    21  Robert Luketic  Jim Sturgess, Kevin Spacey, Kate Bosworth, Air...  United States  January 1, 2020  2008  PG-13  123 min  Dramas  A brilliant group of students become card-count...
```

```
In [5]: netflix.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7787 entries, 0 to 7786
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  --
0   show_id                7787 non-null   object
1   type                  7787 non-null   object
2   title                 7787 non-null   object
3   director              5386 non-null   object
4   cast                  7869 non-null   object
5   country               7288 non-null   object
6   date_added            7777 non-null   object
7   release_year          7787 non-null   int64
8   rating                7780 non-null   object
9   duration              7787 non-null   object
10  listed_in             7787 non-null   object
11  description            7787 non-null   object
dtypes: int64(1), object(11)
memory usage: 730.2+ KB
```

Total movies on netflix

```
In [6]: movies = netflix[netflix['type'] == 'Movie']

In [7]: movies.head(9)

Out[7]:
  show_id  type  title  director  cast  country  date_added  release_year  rating  duration  listed_in  description
1      s2    Movie    7.19  Jorge Michel Grau  Demian Bichir, Héctor Bonilla, Oscar Semán, ...  Mexico  December 23, 2016  2016  TV-MA  93 min  Dramas, International Movies  After a devastating earthquake his Mexico Cit...
2      s3    Movie    23.59  Gilbert Chan  Tedd Chan, Stella Chung, Henley Hii, Lawrence ...  Singapore  December 20, 2011  2011  R  78 min  Horror Movies, International Movies  When an army recruit is found dead, his fellow...
3      s4    Movie    9  Shane Acker  Elijah Wood, John C. Reilly, Jennifer Connelly...  United States  November 16, 2017  2009  PG-13  80 min  Action & Adventure, Independent Movies, Sci-Fi...  In a postapocalyptic world, rag-doll robots hi...
4      s5    Movie    21  Robert Luketic  Jim Sturgess, Kevin Spacey, Kate Bosworth, Air...  United States  January 1, 2020  2008  PG-13  123 min  Dramas  A brilliant group of students become card-count...
6      s7    Movie    122  Yassir Al Yasei  Amina Khalil, Ahmed Dawood, Tarek Lotfy, Ahmed...  Egypt  June 1, 2020  2019  TV-MA  95 min  Horror Movies, International Movies  After an awful accident, a couple admitted to ...
7      s8    Movie    187  Kevin Reynolds  Samuel L. Jackson, John Heard, Kelly Rowland, C...  United States  November 1, 2019  1997  R  119 min  Dramas  After one of his high school students attacks ...
8      s9    Movie    706  Shrawan Kumar  Divya Dutta, Abhi Kulkarni, Mohan Agashe, Anup...  India  April 1, 2019  2019  TV-14  118 min  Horror Movies, International Movies  When a doctor goes missing, his psychiatrist w...
9      s10   Movie    1920  Vikram Bhatt  Rajnesh Duggal, Adah Sharma, Indraneel Sengupta...  India  December 15, 2017  2008  TV-MA  143 min  Horror Movies, International Movies, Thrillers  An architect and his wife move into a castle L...
10     s11   Movie    1922  Zak Hilditch  Thomas Jane, Molly Parker, Dylan Schmid, Kad...  United States  October 20, 2017  2017  TV-MA  103 min  Dramas, Thrillers  A farmer pens a confession admitting to his vi...
```

```
In [8]: total_movies = movies['type'].count()
total_movies

Out[8]: 5377
```

Total TV Shows on netflix

```
In [9]: shows = netflix[netflix['type'] == 'TV Show']

In [10]: shows.head(9)

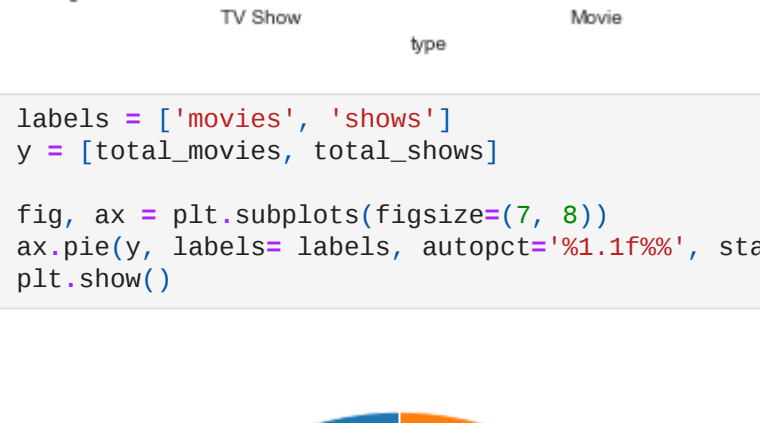
Out[10]:
  show_id  type  title  director  cast  country  date_added  release_year  rating  duration  listed_in  description
0      s1    TV Show    3%      NaN  João Miguel, Bianca Comparato, Michel Gomes, R...  Brazil  August 14, 2020  2020  TV-MA  4 Seasons  International TV Shows, TV Dramas, TV Sci-Fi &...  In a future where the elite inhabit an island ...
5      s6    TV Show    46  Serdar Akar  Erdal Beşikçioğlu, Yasemin Allen, Melis Birkan...  Turkey  July 1, 2017  2016  TV-MA  1 Season  International TV Shows, TV Dramas, TV Mysteries  A genetics professor experiments with a trea...
11     s12    TV Show    1983      NaN  Robert Winkiewicz, Maciej Musiał, Michałina O...  Poland, United States  November 30, 2018  2018  TV-MA  1 Season  Crime TV Shows, International TV Shows, TV Dramas  In this dark alt-history thriller, a novel liv...
12     s13    TV Show    1994  Diego Enrique Osorio      NaN  Mexico  May 17, 2019  2019  TV-MA  1 Season  Crime TV Shows, Documentaries, International TV S...  Archival video and new interviews examine Mexi...
16     s17    TV Show    Feb-09      NaN  Shahd El Yaseen, Shaila Sabi, Hala, Hamed Al...  NaN  March 20, 2019  2018  TV-14  1 Season  International TV Shows, TV Dramas  As a psychology professor faces Alzheimer's h...
24     s25    TV Show    SAINT SEIYA: Knights of the Zodiac      NaN  Bryson Baugus, Emily Nieves, Blake Shepard, Pat...  Japan  January 23, 2020  2020  TV-14  2 Seasons  Anime Series, International TV Shows  Seiya and the Knights of the Zodiac rise again...
26     s27    TV Show    (Un)Well      NaN  NaN  United States  August 12, 2020  2020  TV-MA  1 Season  Reality TV  This docuseries takes a deep dive into the luc...
29     s30    TV Show    #blackAF      NaN  Kenya Baris, Rashida Jones, Imun Benson, Gem...  United States  April 17, 2020  2020  TV-MA  1 Season  TV Comedies  Kenya Baris and his family navigate relation...
38     s39    TV Show    ബഹുമാനപ്പെട്ടനാളിക്ക്      NaN  Chutavuth Pattarakampol, Sheranut Yusananda, N...  NaN  March 30, 2019  2016  TV-14  1 Season  International TV Shows, Romantic TV Shows, TV ...  When two brothers fall for two sisters, they g...
```

```
In [11]: total_shows = shows['type'].count()
total_shows

Out[11]: 2410
```

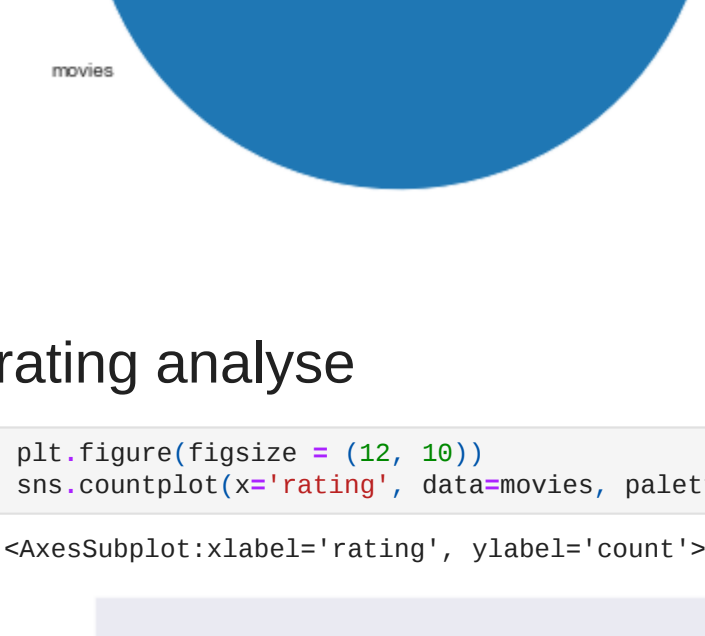
Movies and Shows Visaulation

```
In [12]: sns.set_style('darkgrid')
ax = sns.countplot(x='type', data = netflix, palette = 'Set1')
```



```
In [13]: labels = ['movies', 'shows']
y = [total_movies, total_shows]

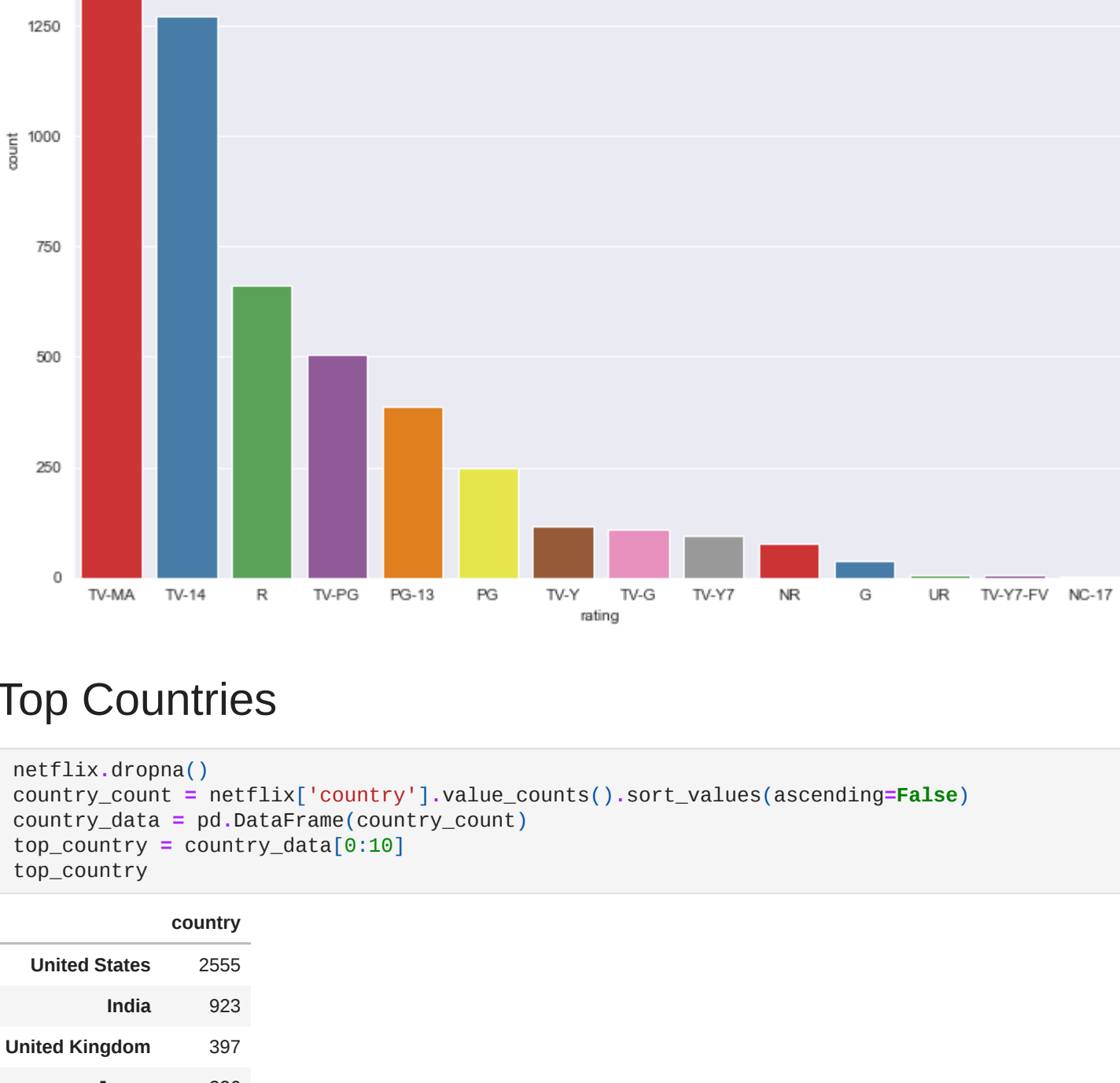
fig, ax = plt.subplots(figsize=(7, 8))
ax.pie(y, labels=labels, autopct='%1.1f%%', startangle=90)
plt.show()
```



rating analyse

```
In [14]: plt.figure(figsize=(12, 10))
sns.countplot(x='rating', data=movies, palette='Set1', order=movies['rating'].value_counts().index[0:15])
```

```
Out[14]: <AxesSubplot: xlabel='rating', ylabel='count'>
```

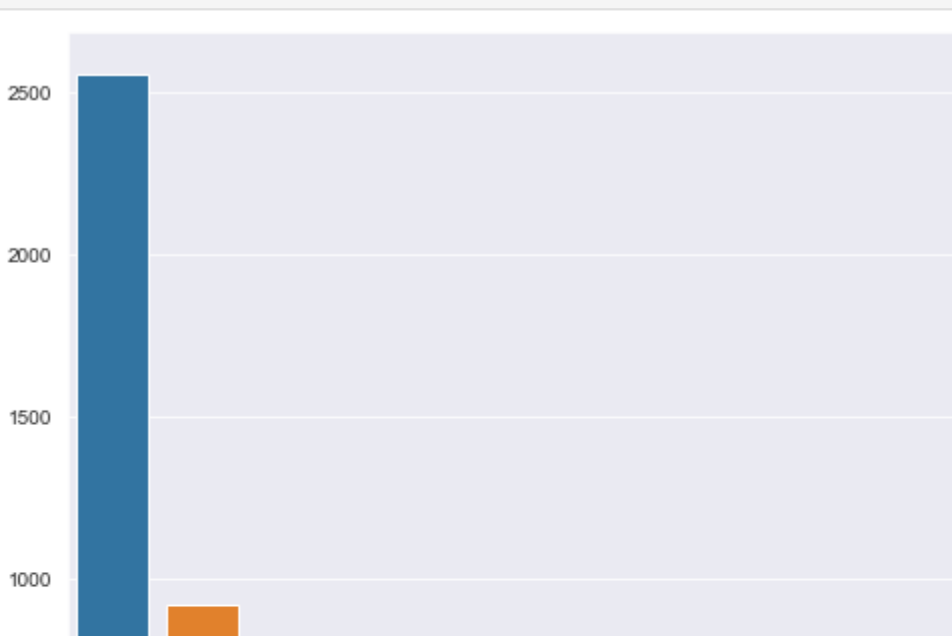


Top Countries

```
In [15]: netflix.dropna()
country_count = netflix['country'].value_counts().sort_values(ascending=False)
country_data = pd.DataFrame(country_count)
top_country = country_data[0:10]
top_country
```

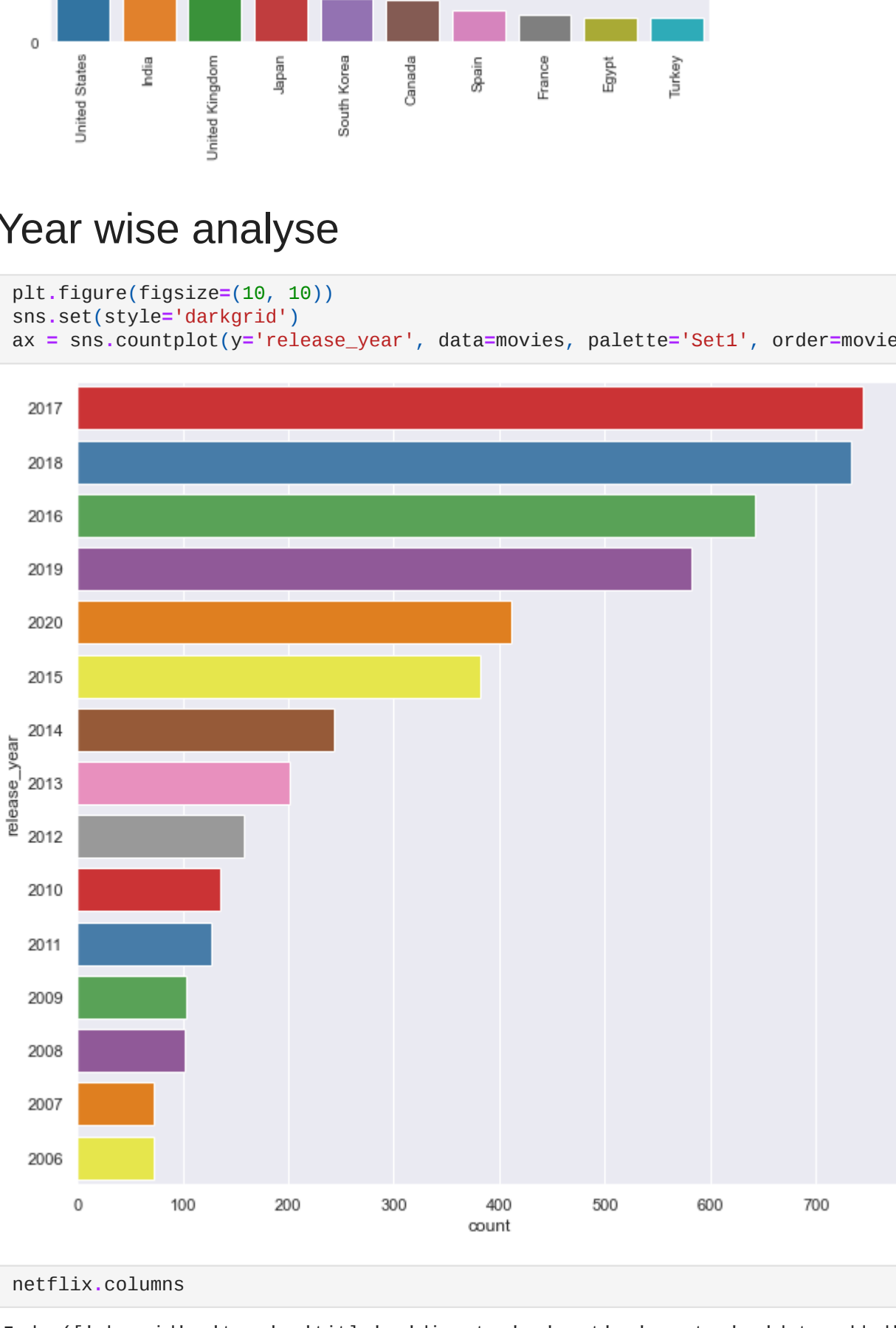
```
Out[15]:
  country
United States  2555
India          923
United Kingdom  397
Japan          226
South Korea    183
Canada         177
Spain          134
France         115
Egypt          101
Turkey         100
```

```
In [16]: plt.figure(figsize=(8, 8))
ax = sns.barplot(x=list(top_country.index), y=np.array(top_country['country']).tolist())
ax.set_xticklabels(list(top_country.index), rotation=90)
plt.show()
```



Year wise analyse

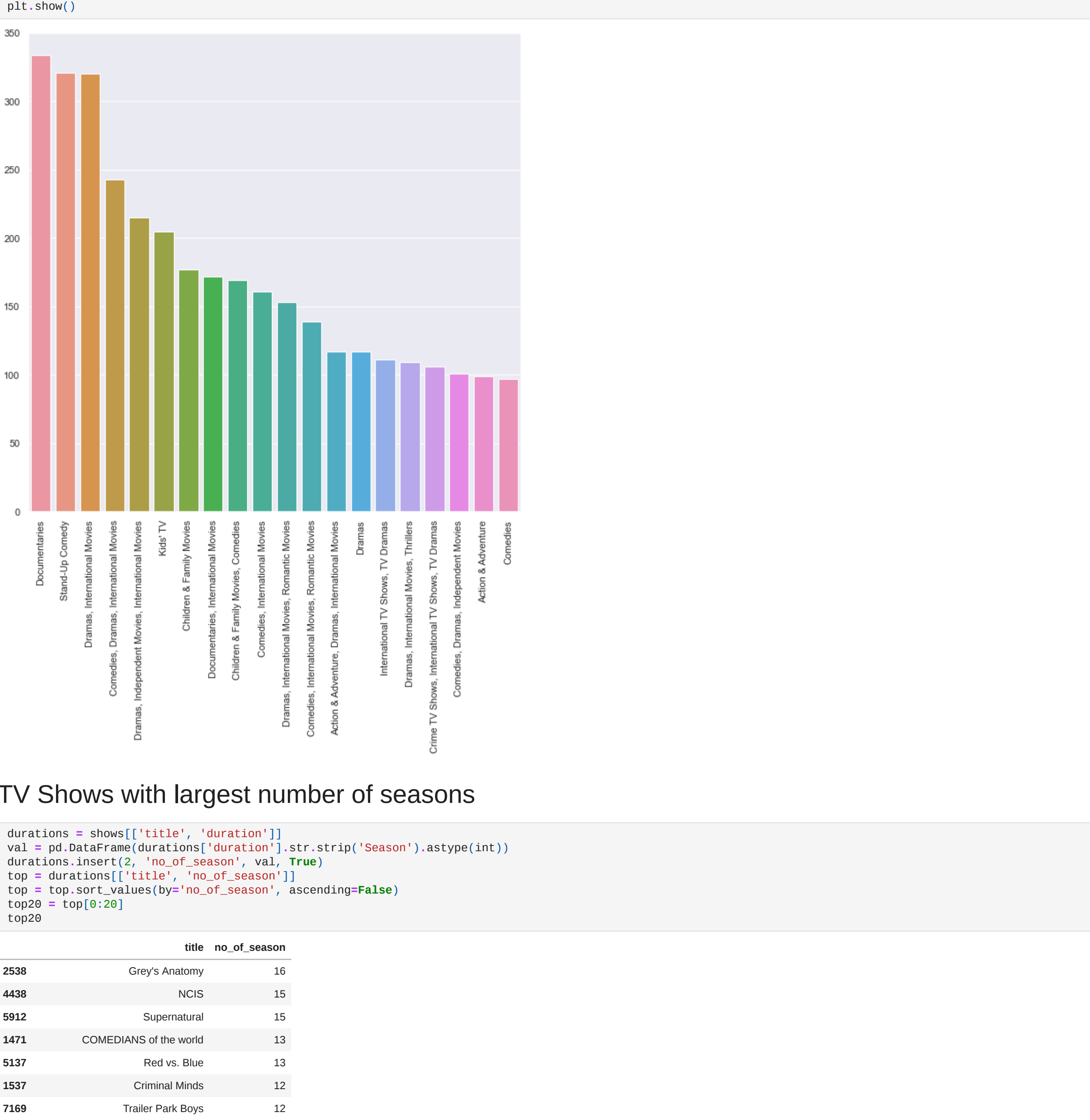
```
In [17]: plt.figure(figsize=(10, 10))
ax = sns.countplot(y='release_year', data=movies, palette='Set1', order=movies['release_year'].value_counts().index[0:15])
```



```
In [18]: netflix.columns

Out[18]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'rating', 'duration', 'listed_in', 'description'], dtype='object')
```

```
In [19]: plt.figure(figsize=(10, 10))
val = sns.barplot(x=netflix['listed_in'].value_counts().sort_values(ascending=False).index[:20], y=netflix['listed_in'].value_counts().sort_values(ascending=False).index[:20], rotation=90)
plt.show()
```



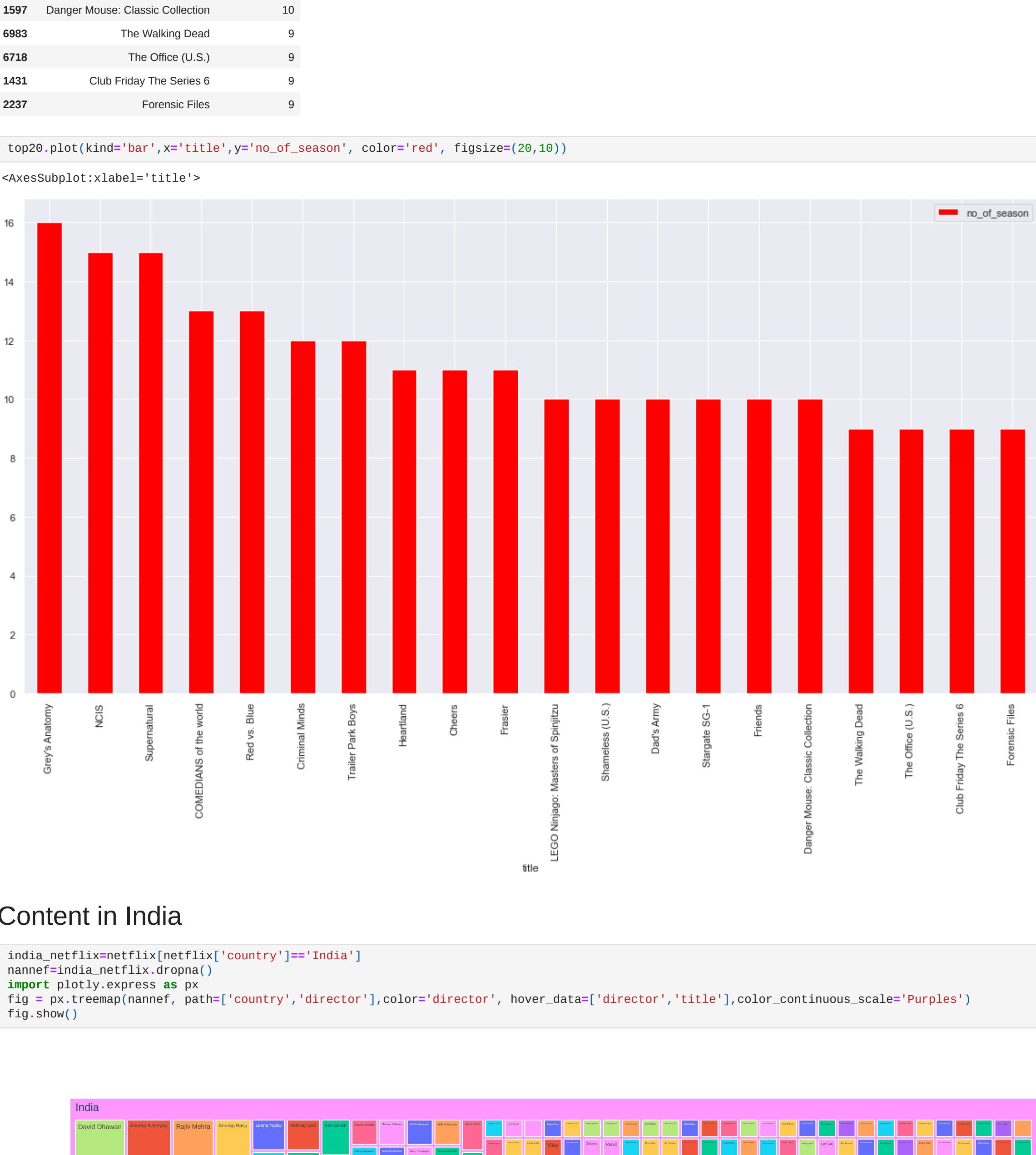
TV Shows with largest number of seasons

```
In [20]: durations = shows[['title', 'duration']]
val = pd.DataFrame(durations['duration'].str.strip('Season').astype(int))
durations.insert(2, 'no_of_season', val, True)
top = durations[['title', 'no_of_season']]
top = top.sort_values(by='no_of_season', ascending=False)
top20 = top[0:20]
top20
```

```
Out[20]:
  title  no_of_season
2538  Grey's Anatomy    16
4438  NCIS             15
5912  Supernatural     15
1471  COMEDIANS of the world  13
5137  Red vs. Blue      13
1537  Criminal Minds   12
7169  Trailer Park Boys  12
2678  Heartland        11
1300  Cheers           11
2263  Frasier           11
3592  LEGO Ninjago: Masters of Spinjitzu  10
5538  Shameless (U.S.)  10
1577  Dad's Army        10
5795  Stargate SG-1     10
2288  Friends           10
1587  Danger Mouse: Classic Collection    10
6983  The Walking Dead   9
6718  The Office (U.S.)  9
1431  Club Friday The Series 6    9
2237  Forensic Files     9
```

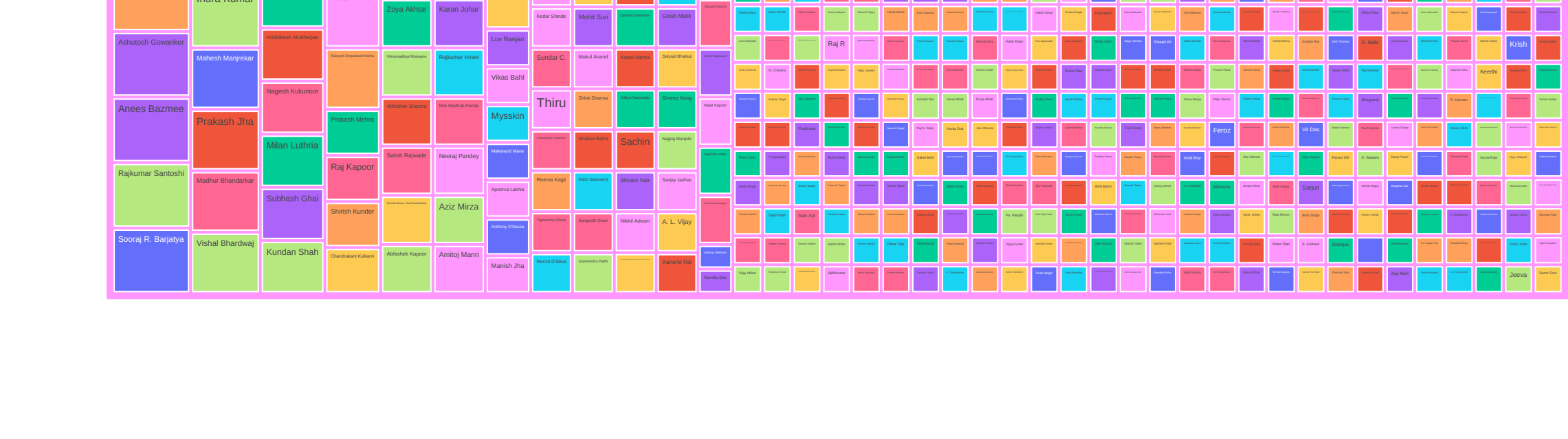
```
In [21]: top20.plot(kind='bar', x='title', y='no_of_season', color='red', figsize=(20, 10))

Out[21]: <AxesSubplot: xlabel='title'>
```



Content in India

```
In [22]: india_netflix=netflix[netflix['country']=='India']
namer=india_netflix.dropna()
import plotly.express as px
fig = px.treemap(namer, path=['country', 'director'], color='director', hover_data=['director', 'title'], color_continuous_scale='Purples')
fig.show()
```



Content based recommendation

```
In [23]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [24]: tfidf = TfidfVectorizer(stop_words='english')
netflix['description'] = netflix['description'].fillna('')
tfidf_matrix = tfidf.fit_transform(netflix['description'])
tfidf_matrix.shape
```

```
Out[24]: (7787, 17905)
```

```
In [25]: from sklearn.metrics.pairwise import linear_kernel
cosine_sim = linear_kernel(tfidf_matrix, tfidf_matrix)
```

```
In [26]: indices = pd.Series(netflix.index, index=netflix['title']).drop_duplicates()
```

```
In [27]: def get_recommendations(title, cosine_sim=cosine_sim):
    idx = indices[title]
    sim_scores = list(enumerate(cosine_sim[idx]))
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)
    sim_scores = sim_scores[1:11]
    movie_indices = [i[0] for i in sim_scores]
    return netflix['title'].iloc[movie_indices]
```

```
In [28]: get_recommendations('Peaky Blinders')
```

```
Out[28]:
4692  Our Godfather
4358  My Stupid Boss
1897  Don
6344  The Fear
3219  Jonathan Strange & Mr Norrell
4963  Power Rangers Zeo
6783  The Prison
6950  The Tudors
6236  The Con Is On
6585  The Legend of Michael Mishra
Name: title, dtype: object
```

```
In [29]: get_recommendations('Mortal')
```

```
Out[29]:
4920  PTLI Fantasy: War of Dragons
1858  Edgar Rice Burroughs' Tarzan and Jane
2159  Figures of Speech
4314  My Entire High School Sinking Into the Sea
2188  FirstBorn
4904  Psychokinesis
608  At First Light
6958  The Umbrella Academy
1257  Channelker
3864  Maharakshak Devi
Name: title, dtype: object
```