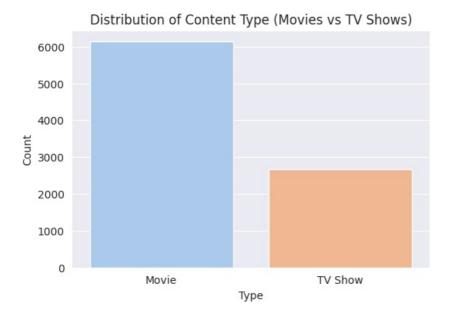
```
import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
         Importing Dataset
In [2]: df = pd.read csv('/kaggle/input/exploratory-data-analysis-on-netflix-data/netflix titles 2021.csv')
         Data head
In [3]:
        df.head()
Out[3]:
                                                                                                                    listed_in description
            show_id
                       type
                                  title
                                       director
                                                      cast country date_added release_year rating duration
                                                                                                                                   As her
                                  Dick
                                                                                                                              father nears
                                         Kirsten
                                                             United
                                                                      September
                                                                                                PG-
         0
                  s1 Movie
                               Johnson
                                                      NaN
                                                                                        2020
                                                                                                        90 min Documentaries
                                                                                                                                the end of
                                                             States
                                                                        25, 2021
                                        Johnson
                               Is Dead
                                                                                                                                  his life,
                                                                                                                                  filmm...
                                                      Ama
                                                                                                                                    After
                                                  Qamata,
                                                                                                                 International
                                                                                                                                 crossing
                                                     Khosi
                        TV
                                                                                                 TV-
                                                                                                            2
                               Blood &
                                                              South
                                                                      September
                                                                                                                TV Shows, TV
                                                                                                                                paths at a
                                           NaN
                                                   Ngema,
                                                                                        2021
                      Show
                                 Water
                                                                        24, 2021
                                                                                                      Seasons
                                                                                                                  Dramas, TV
                                                              Africa
                                                                                                 MA
                                                                                                                                 party, a
                                                      Gail
                                                                                                                    Mysteries
                                                                                                                              Cape Town
                                                 Mabalane,
                                                                                                                                      t...
                                                  Thaban...
                                                     Sami
                                                  Bouajila,
                                                                                                                    Crime TV
                                                                                                                                To protect
                                                     Tracy
                                                                                                                      Shows,
                                                                                                                                his family
                                          Julien.
                                                                                                 TV-
                        T\
                                                                      September
                                                                                                            1
                                                                                        2021
                             Ganglands
                                                    Gotoas,
                                                               NaN
                                                                                                                 International
                                                                                                                                   from a
                      Show
                                        Leclercq
                                                                                                 MA
                                                                        24. 2021
                                                                                                       Season
                                                                                                                TV Shows, TV
                                                                                                                                 powerful
                                                    Samuel
                                                     Jouy,
                                                                                                                                drug lor...
                                                                                                                        Act
                                                    Nabi...
                                                                                                                                  Feuds,
                                                                                                                                 flirtations
                               Jailbirds
                        TV
                                                                                                 TV-
                                                                                                                  Docuseries.
                                                                                                                                and toilet
                                                                      September
                                                                                                            1
         3
                                  New
                                           NaN
                                                      NaN
                                                               NaN
                                                                                        2021
                      Show
                                                                        24, 2021
                                                                                                 MA
                                                                                                       Season
                                                                                                                                  talk go
                                                                                                                   Reality TV
                               Orleans
                                                                                                                                   down
                                                                                                                                   amo...
                                                    Mayur
                                                     More,
                                                                                                                                In a city of
                                                                                                                 International
                                                   Jitendra
                                                                                                                                 coaching
                        TV
                                                                      September
                                                                                                 TV-
                                                                                                            2
                                                                                                                   TV Shows,
                                  Kota
                                           NaN
                                                               India
                                                                                        2021
                  s5
                                                    Kumar,
                                                                                                                                  centers
                                                                        24, 2021
                      Show
                                Factory
                                                                                                 MA
                                                                                                      Seasons
                                                                                                                 Romantic TV
                                                    Ranian
                                                                                                                                 known to
                                                                                                                 Shows, TV ...
                                                  Raj, Alam
                                                                                                                                  train I...
                                                       K...
                                                                                                                                     -
In [4]: df.shape
Out[4]: (8807, 12)
         Information Dataset
In [5]: df.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 8807 entries, 0 to 8806
        Data columns (total 12 columns):
                             Non-Null Count Dtype
        # Column
        - - -
             -----
                              ______
         0
             show id
                             8807 non-null
                                                obiect
         1
             type
                             8807 non-null
                                                object
         2
             title
                             8807 non-null
                                                object
         3
             director
                             6173 non-null
                                                object
                             7982 non-null
         4
             cast
                                                object
             country
         5
                             7976 non-null
                                                object
         6
             date added
                             8797 non-null
                                                object
         7
             release_year
                             8807 non-null
                                                int64
         8
             rating
                             8803 non-null
                                                object
         9
             duration
                             8804 non-null
                                                object
         10
            listed in
                             8807 non-null
                                                object
         11 description
                             8807 non-null
                                                object
        dtypes: int64(1), object(11)
        memory usage: 825.8+ KB
         Duplicated Values
```

In [1]: import pandas as pd

In [6]: df.duplicated().sum()

Dropping missing values

```
In [7]: | df.drop(['show id', 'cast', 'date added', 'description'], inplace=True, axis=1)
           df.head()
                                    title
                                               director
                                                           country release_year rating
                                                                                         duration
                                                                                                                                  listed_in
 Out[7]:
                 type
                                                Kirsten
                                                             United
                          Dick Johnson Is
                                                                                    PG-
           0
                Movie
                                                                            2020
                                                                                           90 min
                                                                                                                            Documentaries
                                                                                     13
                                   Dead
                                               Johnson
                                                             States
                  TV
                                                             South
                                                                                    TV-
                                                                                                2
                                                                                                       International TV Shows, TV Dramas, TV
                           Blood & Water
                                                  NaN
                                                                            2021
                Show
                                                                                         Seasons
                                                             Africa
                                                                                    MA
                                                                                                                                 Mysteries
                                                 Julien
                                                                                    TV-
                                                                                                     Crime TV Shows, International TV Shows,
                  TV
           2
                              Ganglands
                                                                           2021
                                                              NaN
                Show
                                              Leclercq
                                                                                    MA
                                                                                          Season
                                                                                                                                  TV Act...
                  TV
                            Jailbirds New
                                                                                    TV-
           3
                                                  NaN
                                                              NaN
                                                                           2021
                                                                                                                      Docuseries, Reality TV
                Show
                                 Orleans
                                                                                    MA
                                                                                          Season
                  TV
                                                                                    TV-
                                                                                                2
                                                                                                         International TV Shows, Romantic TV
           4
                             Kota Factory
                                                  NaN
                                                              India
                                                                           2021
                Show
                                                                                    MA
                                                                                         Seasons
                                                                                                                              Shows, TV ...
 In [8]: df.fillna('Unknown',inplace=True,axis=1)
 In [9]: df.isnull().sum()
 Out[9]: type
           title
                              0
                              0
           director
           country
                              0
           release_year
                              0
           rating
                              0
           duration
                              0
           listed in
                              0
           dtype: int64
           Exploratory Data
In [10]: df.head()
Out[10]:
                                    title
                                               director
                                                           country release_year rating
                                                                                         duration
                                                                                                                                  listed_in
                 type
                          Dick Johnson Is
                                                Kirsten
                                                             United
                                                                                    PG-
           0
                Movie
                                                                           2020
                                                                                           90 min
                                                                                                                            Documentaries
                                               Johnson
                                                             States
                                                                                     13
                                   Dead
                  TV
                                                             South
                                                                                    TV-
                                                                                                2
                                                                                                       International TV Shows, TV Dramas, TV
                           Blood & Water
                                              Unknown
                                                                           2021
                Show
                                                             Africa
                                                                                    MA
                                                                                         Seasons
                                                 Julien
                                                                                                     Crime TV Shows, International TV Shows,
                  T\/
                                                                                    TV-
                                                                                                1
           2
                              Ganglands
                                                          Unknown
                                                                           2021
                Show
                                              Leclercq
                                                                                    MA
                                                                                          Season
                                                                                                                                  TV Act...
                  ΤV
                            Jailbirds New
                                                                                    TV-
           3
                                              Unknown
                                                          Unknown
                                                                            2021
                                                                                                                      Docuseries, Reality TV
                Show
                                                                                    MA
                                 Orleans
                                                                                          Season
                  TV
                                                                                    TV-
                                                                                                2
                                                                                                         International TV Shows, Romantic TV
           4
                             Kota Factory
                                              Unknown
                                                              India
                                                                           2021
                Show
                                                                                    MA
                                                                                         Seasons
                                                                                                                              Shows. TV ...
           Data Visualization
In [11]: sns.set_style("darkgrid")
           plt.figure(figsize=(6, 4))
           sns.countplot(x=df['type'], palette="pastel")
           plt.title("Distribution of Content Type (Movies vs TV Shows)")
           plt.xlabel("Type")
           plt.ylabel("Count")
           plt.show()
```

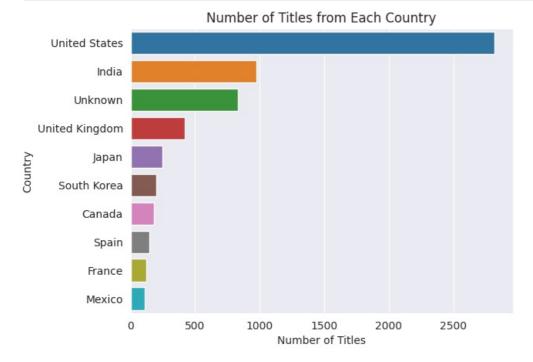


Add a new column with the new category names

In [12]:	df	<pre>n_largest_titles = df['country'].value_counts().nlargest(10).index.tolist() df['country_filtered'] = df['country'].apply(lambda x: x if x in n_largest_titles else "Other") df.head()</pre>									
Out[12]:		type	title	director	country	release_year	rating	duration	listed_in	country_filtered	
	0	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2020	PG- 13	90 min	Documentaries	United States	
	1	TV Show	Blood & Water	Unknown	South Africa	2021	TV- MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	Other	
	2	TV Show	Ganglands	Julien Leclercq	Unknown	2021	TV- MA	1 Season	Crime TV Shows, International TV Shows, TV Act	Unknown	
	3	TV Show	Jailbirds New Orleans	Unknown	Unknown	2021	TV- MA	1 Season	Docuseries, Reality TV	Unknown	
	4	TV Show	Kota Factory	Unknown	India	2021	TV- MA	2 Seasons	International TV Shows, Romantic TV Shows, TV	India	

Plot the chart

```
In [13]: ax = sns.countplot(y='country_filtered', data=df.dropna(subset='country'), order=n_largest_titles)
    ax.set_xlabel('Number of Titles')
    ax.set_ylabel('Country')
    ax.set_title('Number of Titles from Each Country');
```

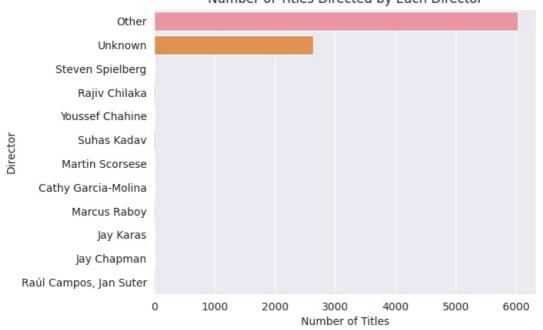


How many movies did each director direct?

Just like above, we'll only plot the top directors.

```
In [14]:
          director value counts = df['director'].value counts()
           director_value_counts[director_value_counts >= 10]
           # there are eleven directors with more than 10 titles so we use those
           n_largest_directors = director_value_counts.nlargest(11).index
           n_largest_directors
Out[14]: Index(['Unknown', 'Rajiv Chilaka', 'Raúl Campos, Jan Suter', 'Suhas Kadav', 'Marcus Raboy', 'Jay Karas', 'Cathy Garcia-Molina', 'Jay Chapman',
                   'Youssef Chahine', 'Martin Scorsese', 'Steven Spielberg'],
                  dtype='object', name='director')
In [15]: df['director filtered'] = df['director'].apply(lambda x: x if x in n largest directors else "Other")
           df.head()
Out[15]:
               type
                            title
                                   director
                                             country release_year
                                                                   rating duration
                                                                                               listed_in country_filtered director_filtered
                            Dick
                                    Kirsten
                                              United
                                                                     PG-
           0 Movie
                      Johnson Is
                                                             2020
                                                                             90 min
                                                                                          Documentaries
                                                                                                            United States
                                                                                                                                   Other
                                                                       13
                                   Johnson
                                              States
                           Dead
                                                                                         International TV
                \mathsf{TV}
                         Blood &
                                               South
                                                                                 2
                                  Unknown
                                                             2021
                                                                                      Shows, TV Dramas,
                                                                                                                  Other
                                                                                                                                Unknown
              Show
                           Water
                                               Africa
                                                                      MA
                                                                           Seasons
                                                                                           TV Mysteries
                                                                                        Crime TV Shows,
                                                                      TV-
                TV
                                     Julien
                      Ganglands
                                            Unknown
                                                             2021
                                                                                         International TV
                                                                                                               Unknown
                                                                                                                                   Other
              Show
                                   Leclercq
                                                                      MA
                                                                            Season
                                                                                        Shows, TV Act...
                         Jailbirds
                                                                      TV-
                TV
                                                                                      Docuseries, Reality
                           New
                                  Unknown Unknown
                                                             2021
                                                                                                               Unknown
                                                                                                                                Unknown
              Show
                                                                      MA
                                                                            Season
                         Orleans
                                                                                         International TV
                                                                                  2
                            Kota
                                  Unknown
                                                             2021
                                                                                                                   India
                                                                                                                                Unknown
                                                India
                                                                                     Shows, Romantic TV
              Show
                                                                      MA
                                                                           Seasons
                         Factory
                                                                                           Shows, TV ...
In [16]: ax = sns.countplot(y='director_filtered', data=df.dropna(subset='director'))
           ax.set_xlabel('Number of Titles')
           ax.set ylabel('Director')
           ax.set_title('Number of Titles Directed by Each Director');
```

Number of Titles Directed by Each Director



In [17]: df.head()

Out[17]:		type	title	director	country r	elease_year	rating	dura	tion	listed_in	country_filtered	director_filtered
	0	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2020	PG- 13	90	min	Documentaries	United States	Other
	1	TV Show	Blood & Water	Unknown	South Africa	2021	TV- MA		2 sons	International TV Shows, TV Dramas, TV Mysteries	Other	Unknown
	2	TV Show	Ganglands	Julien Leclercq	Unknown	2021	TV- MA		1 ason	Crime TV Shows, International TV Shows, TV Act	Unknown	Other
	3	TV Show	Jailbirds New Orleans	Unknown	Unknown	2021	TV- MA		1 ason	Docuseries, Reality TV	Unknown	Unknown
	4	TV Show	Kota Factory	Unknown	India	2021	TV- MA		2 sons S	International TV Shows, Romantic TV Shows, TV	India	Unknown
In [18]:		_	s = df.loc[s.head()	df['type']	== 'Movi	e'].copy()	.reset	_inde	x(drop	=True)		
	uı		3 i licau ()									
Out[18]:	uı	type	title	directo	r count	try release_	year r	ating	duratio	on listed_in	country_filtered	director_filtered
Out[18]:	0	_		directo Kirster Johnson	n Unit	ed ,	year r	PG- 13	duratic 90 m		country_filtered United States	director_filtered Other
Out[18]:	0	type	title Dick Johnson Is	Kirster	n Unit n Stat t é Unknor	ed ,		PG-		nin Documentaries	<u> </u>	
Out[18]:	0	type Movie	Dick Johnson Is Dead My Little Pony: A New	Kirster Johnson Rober Cullen, Jose	t Unknor	ed es des des des des des des des des de	2020	PG- 13	90 m	nin Documentaries Children & Family Movies Dramas, Independent	United States	Other
Out[18]:	0	type Movie Movie	Dick Johnson Is Dead My Little Pony: A New Generation	Kirster Johnson Rober Cullen, Jose Luis Ucha	t Unknor	ed ees ees ees ees ees ees ees ees ees e	2020	PG- 13 PG	90 m	nin Documentaries Children & Family Movies Dramas, Independent Movies, International Movies Comedica	United States Unknown	Other

Checking if all the rows have the same unit (min)

```
In [19]: df_movies.duration.str.contains('min').sum() == df_movies.dropna(subset='duration').shape[0]
Out[19]: False
```

Converting the durations to numbers

df movies['duration'] = pd.to numeric(df movies['duration'].apply(lambda x: str(x).split()[0]), errors='coerce' df movies.head() Out[20]: director country release_year rating duration listed_in country_filtered director_filtered type Dick PG-Kirsten United 0 Movie 2020 Documentaries **United States** Other Johnson Is Johnson States Dead My Little Robert Pony: A Children & Cullen, José 2021 PG 91.0 Other 1 Unknown Unknown Movie **Family Movies** New Luis Ucha Generation United Dramas. States, Independent Haile Ghana, 1993 125.0 Movies, Other Other 2 Movie Sankofa Burkina Gerima International Faso, Movies United Kin... Theodore United PG-Comedies, Movie The Starling 2021 104.0 **United States** Other Melfi States Dramas Germany, Dramas, Christian TV-4 Movie Je Suis Karl Czech 2021 127.0 International Other Other Schwochow MA Republic

Create a new boolean column

```
In [21]: df_movies['longer_than_90'] = df_movies['duration'] > 90

/usr/local/lib/python3.11/dist-packages/pandas/core/computation/expressions.py:73: RuntimeWarning: invalid value encountered in greater return op(a, b)
```

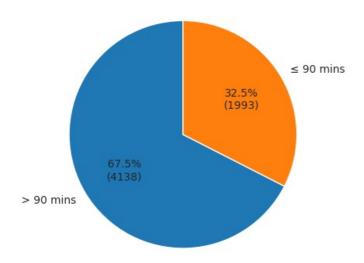
Creating a function so that the chart shows the respective

```
In [22]:
    def movie_autopct(pct):
        total = sum(df_movies['longer_than_90'].value_counts())
        count = int(round(pct * total / 100.0))
        return f'{pct:.1f}%\n({count})'

# Plotting the chart

df_movies['longer_than_90'].value_counts().plot.pie(
        labels=['> 90 mins', '≤ 90 mins'],
        autopct=movie_autopct,
        startangle=90,
        ylabel=''
    )
    plt.title('Proportions of Movie Lengths')
    plt.show()
```

Proportions of Movie Lengths



In [23]:	df	.head()									
Out[23]:		type	title	director	country	release_year	rating	duration	listed_in	country_filtered	director_filtered
	0	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2020	PG- 13	90 min	Documentaries	United States	Other
	1	TV Show	Blood & Water	Unknown	South Africa	2021	TV- MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	Other	Unknown
	2	TV Show	Ganglands	Julien Leclercq	Unknown	2021	TV- MA	1 Season	Crime TV Shows, International TV Shows, TV Act	Unknown	Other
	3	TV Show	Jailbirds New Orleans	Unknown	Unknown	2021	TV- MA	1 Season	Docuseries, Reality TV	Unknown	Unknown
	4	TV Show	Kota Factory	Unknown	India	2021	TV- MA	2 Seasons	International TV Shows, Romantic TV Shows, TV	India	Unknown

What is the shortest movie?

```
In [24]: df_movies.sort_values(by='duration').head(10)
```

Out[24]:		type	title	director	country	release	e_year	rating	duration	listed_in	country_filtered	director_filtered
	2402	Movie	Silen	Limbert Fabian, Brandon Oldenburg	United		2014	TV-Y	3.0	Children & Family Movies, Sci-Fi & Fantasy	United States	Other
	1777	Movie	Sol Levante	Akira Saitoh			2020	TV-14	5.0	Action & Adventure, Anime Features, Internatio	Japar	Other
	972	Movie	Cops and Robbers		United States		2020	PG- 13	8.0	Dramas	United States	Other
	1017	Movie	Canvas	Frank E. Abney III	United States		2020	G	9.0	Children & Family Movies, Dramas	United States	Other
	2276	Movie	Americar Factory: A Conversatior with the Obamas	u Unknown	United States		2019	TV-G	10.0	Documentaries	United States	Unknown
	1873	Movie	Calico Critters Everyone's Big Dream Flying i	Momoko Kamiya	Linknown		2019	TV-Y	11.0	Children & Family Movies	Unknowr	Other
	4229	Movie	Calico Critters: A Town o Dreams	Momoko f Kamiya	LINKNOWN		2017	TV-Y	11.0	Children & Family Movies	Unknowr	Other
	2400	Movie	Cosmos Laundromat First Cycle	Mathieu			2015	TV- MA	12.0	Dramas, International Movies, Sci-Fi & Fantasy	Other	Other
	3008	Movie	Zior	Floyd Russ			2018	TV- PG	12.0	Documentaries, Sports Movies	United States	Other
	439	Movie	Besieged Bread		Linknown		2015	TV-14	12.0	Dramas, International Movies	Unknowr	Other
	4	1())
In [25]: Out[25]:	df.h	ead ()	title	director co	ountry releas	e year	rating	duratio	on	listed in	country_filtered	director filtered
	0 M	ovie	Dick Johnson Is Dead		United States	2020	PG- 13	90 m	nin	Documentaries	United States	Other
	1 _S	TV how	Blood & Water	Inknown	South Africa	2021	TV- MA	Seaso	2 Show	nternational TV vs, TV Dramas, TV Mysteries	Other	Unknown
	2 S	TV how	Ganglands	Julien Leclercq Un	known	2021	TV- MA	Seas	on I	ime TV Shows, nternational TV hows, TV Act	Unknown	Other
	3 _S	TV how	Jailbirds New U Orleans	Inknown Un	known	2021	TV- MA	Seas		useries, Reality TV	Unknown	Unknown
	4 S	TV how	Kota Factory	nknown	India	2021	TV- MA	Seaso	2 Show	nternational TV s, Romantic TV Shows, TV	India	Unknown

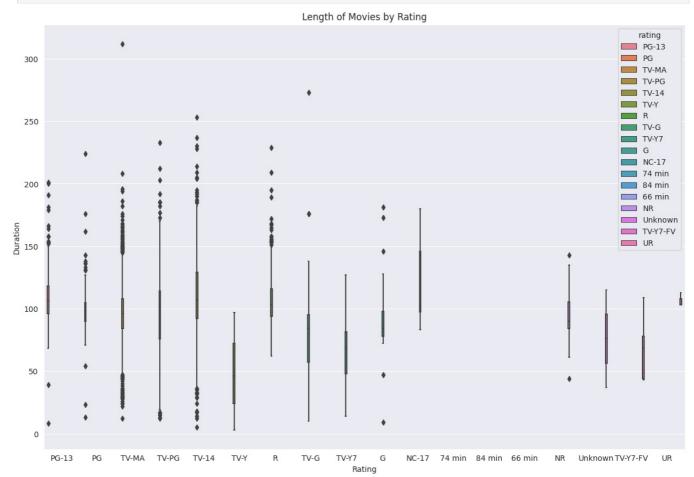
In [26]: df_movies.head()

t[26]:		type	title	director	country	release_year	rating	duration	listed_in	country_filtered	director_filtered	longer
	0	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2020	PG- 13	90.0	Documentaries	United States	Other	
	1	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Unknown	2021	PG	91.0	Children & Family Movies	Unknown	Other	
	2	Movie	Sankofa	Haile Gerima	United States, Ghana, Burkina Faso, United Kin	1993	TV- MA	125.0	Dramas, Independent Movies, International Movies	Other	Other	
	3	Movie	The Starling	Theodore Melfi	United States	2021	PG- 13	104.0	Comedies, Dramas	United States	Other	
	4	Movie	Je Suis Karl	Christian Schwochow	Germany, Czech Republic	2021	TV- MA	127.0	Dramas, International Movies	Other	Other	
	4											b

Length of Movies

A boxplot seems appropriate to me in this scenario. It can show how the ratings relate to the length of the movies, while giving us valuable insights regarding the outliers.

```
In [27]:
    plt.figure(figsize=(15, 10))
    sns.boxplot(x='rating', y='duration', data=df_movies, hue='rating', palette='husl')
    plt.xlabel('Rating')
    plt.ylabel('Duration')
    plt.title('Length of Movies by Rating');
```



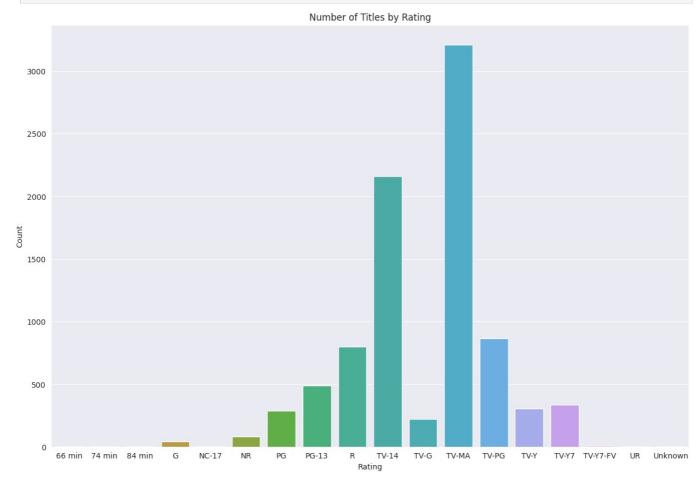
get the count and rating into a dataframe

```
rating_count_ser = df.groupby('rating').duration.count()
rating_count_df = rating_count_ser.to_frame().reset_index()
rating_count_df.columns=['rating', 'duration_count']
rating_count_df
```

	rating	duration_count
0	66 min	1
1	74 min	1
2	84 min	1
3	G	41
4	NC-17	3
5	NR	80
6	PG	287
7	PG-13	490
8	R	799
9	TV-14	2160
10	TV-G	220
11	TV-MA	3207
12	TV-PG	863
13	TV-Y	307
14	TV-Y7	334
15	TV-Y7-FV	6
16	UR	3
17	Unknown	4

Out[28]:

```
In [29]:
    plt.figure(figsize=(15, 10))
    sns.barplot(data=rating_count_df, x='rating', y='duration_count')
    plt.xlabel('Rating')
    plt.ylabel('Count')
    plt.title('Number of Titles by Rating');
```



Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js