In [3]:
```
import pandas as pd
```

In [4]:
```
dataset=pd.read_csv('netflix_titles.csv',index_col='show_id')
dataset
```

Out[4]:

show_id	type	title	director	cast	country	date_added	release_year	rating	duratic
s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 m
s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	Seasor
s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	Seasc
s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	Seasc
s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	Se
...	...	...	...	...	...	...	...	...	
s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 m
s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	Seasor
s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 m
s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 m

show_id	type	title	director	cast	country	date_added	release_year	rating	duratic
s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 m

8807 rows × 11 columns

In [5]:
```
dataset.drop(['description','cast','director'],axis=1,inplace=True)
```

In [6]:
```
dataset
```

Out[6]:

show_id	type	title	country	date_added	release_year	rating	duration	listed_in
s1	Movie	Dick Johnson Is Dead	United States	September 25, 2021	2020	PG-13	90 min	Documentaries
s2	TV Show	Blood & Water	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries
s3	TV Show	Ganglands	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
s4	TV Show	Jailbirds New Orleans	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
s5	TV Show	Kota Factory	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...
...	...	...	...	...	...	...	...	...
s8803	Movie	Zodiac	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers
s8804	TV Show	Zombie Dumb	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Kore TV Shows, T Comedie
s8805	Movie	Zombieland	United States	November 1, 2019	2009	R	88 min	Comed Horror Movies
s8806	Movie	Zoom	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies
s8807	Movie	Zubaan	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals

8807 rows × 8 columns

In [7]:
```
dataset.columns
```

Out[7]:
```
Index(['type', 'title', 'country', 'date_added', 'release_year', 'rating',
       'duration', 'listed_in'],
      dtype='object')
```

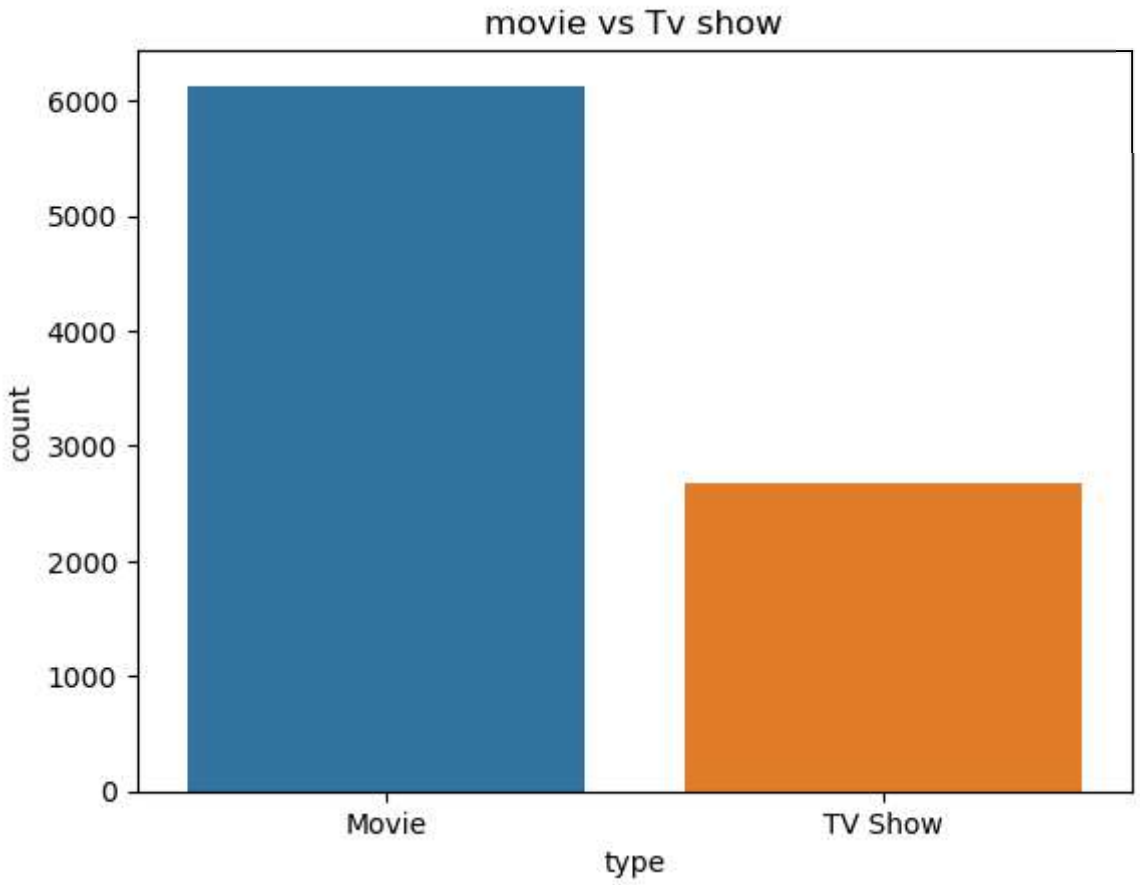
In [8]:
```
missing_values=dataset.isnull().sum()
missing_values
```

Out[8]:
```
type             0
title            0
country        831
date_added      10
release_year     0
rating           4
duration         3
listed_in        0
dtype: int64
```

In [9]:
```
dataset['country'].fillna('Unknown', inplace=True)
dataset['date_added'].fillna('Unknown', inplace=True)
dataset['duration'].fillna('Unknown', inplace=True)
dataset['rating'].fillna('Unknown', inplace=True)
```

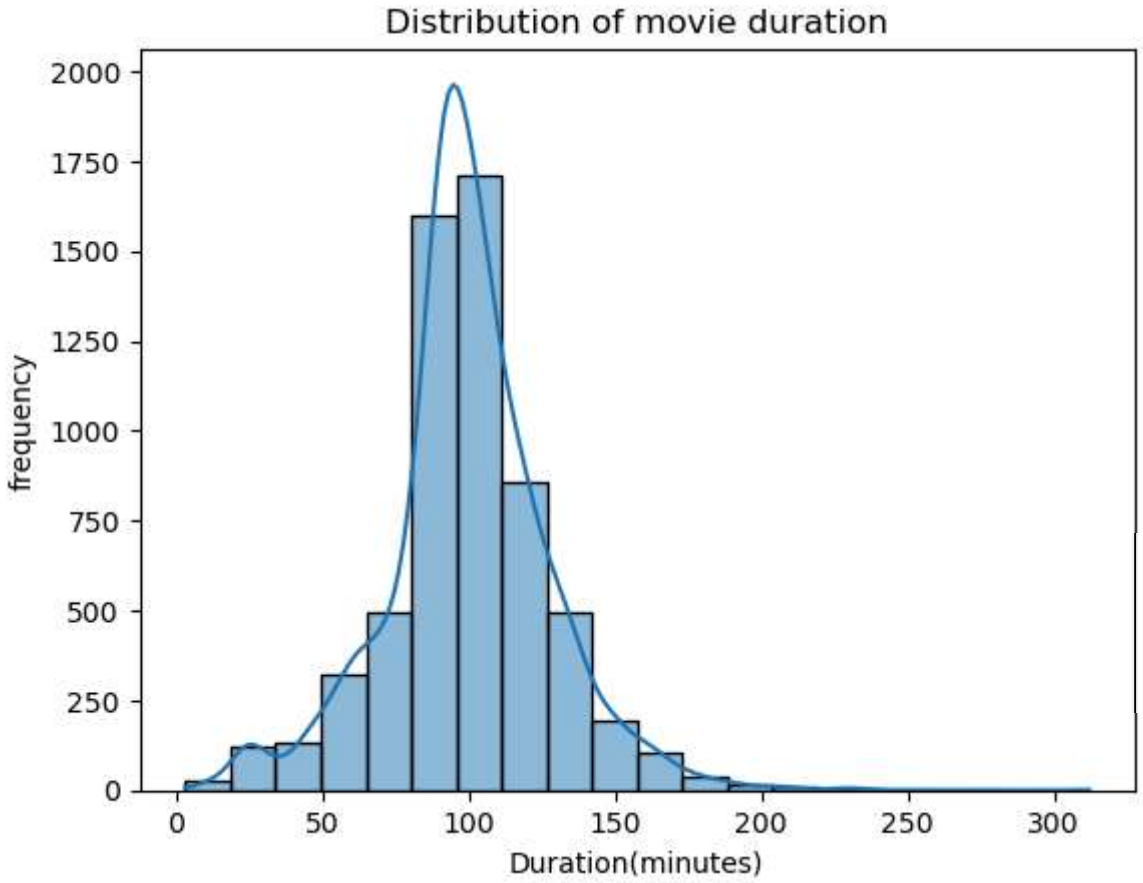
In [39]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

In [40]:
```
sns.countplot(data=dataset,x='type')
plt.title('movie vs Tv show')
plt.show()
```

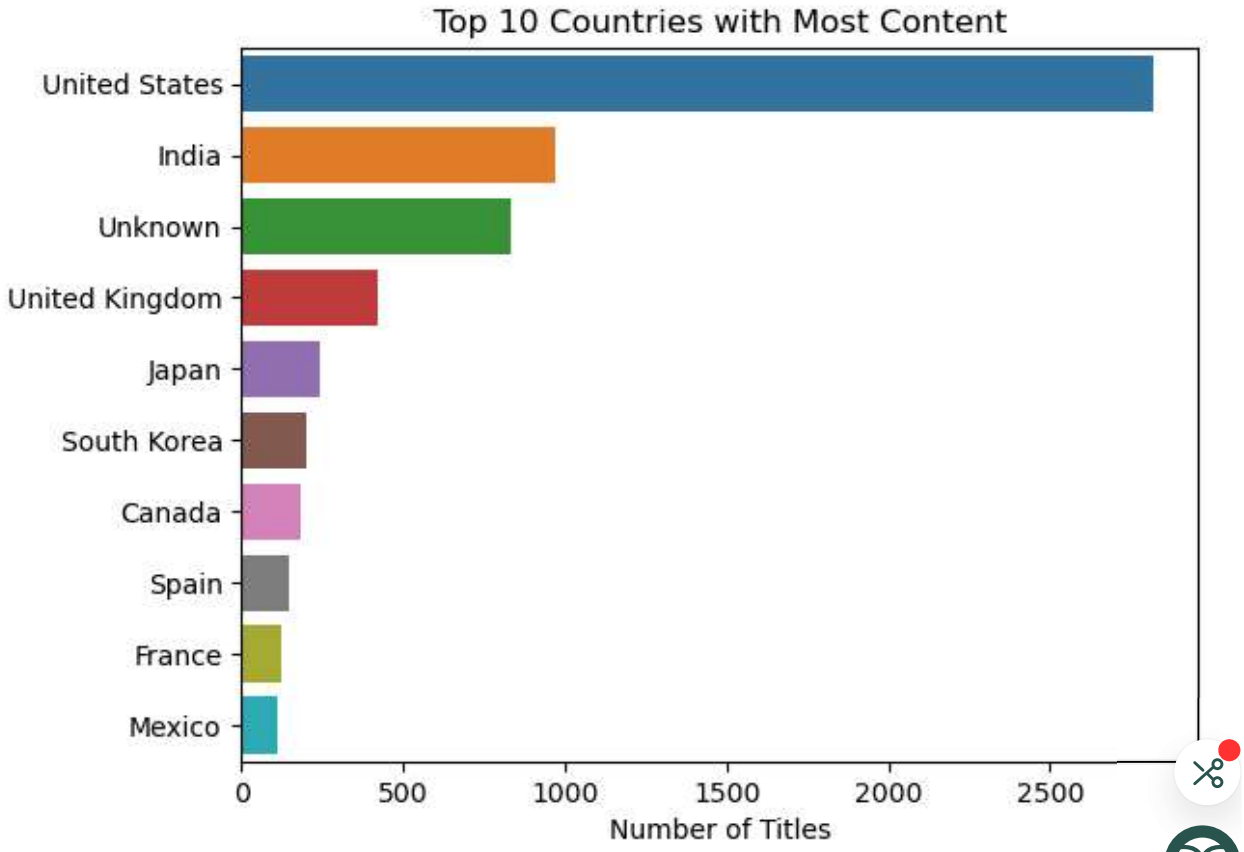


In [41]:
```
df=dataset[dataset['duration'].str.contains('min',na=False)].copy()
df['duration_mintues']=df['duration'].str.replace('min','').astype(int)
```

In [42]:
```
sns.histplot(df['duration_mintues'],bins=20,kde=True)
plt.title('Distribution of movie duration')
plt.xlabel('Duration(minutes)')
plt.ylabel('frequency')
plt.show()
```

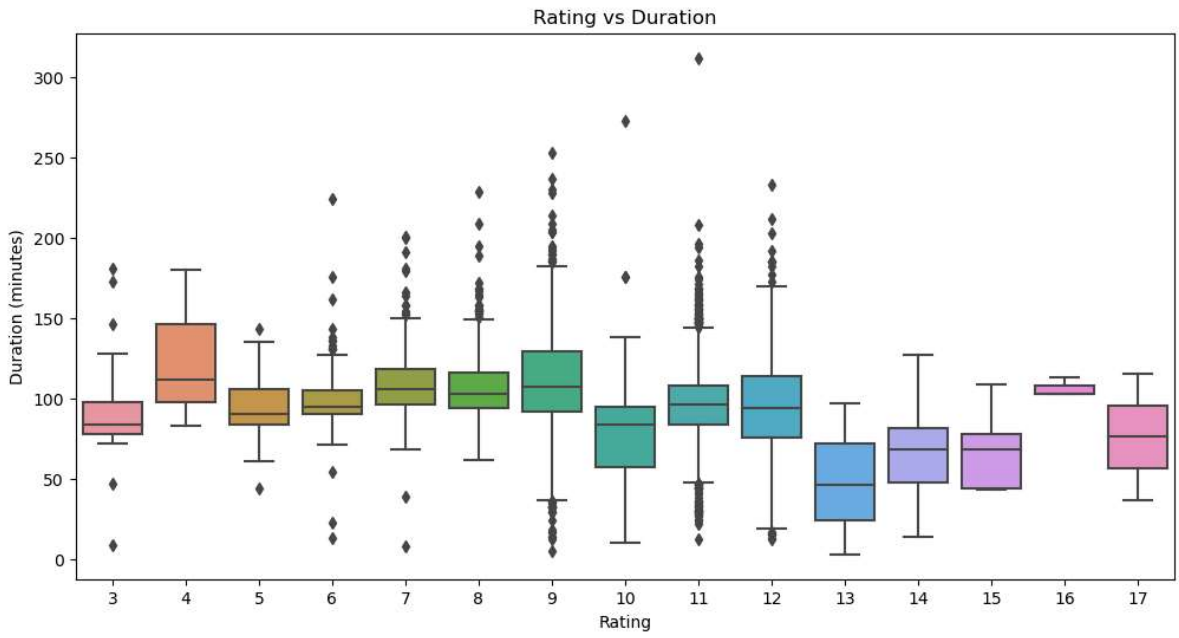


In [43]:
```
top_countries = dataset['country'].value_counts().head(10)
sns.barplot(x=top_countries.values, y=top_countries.index)
plt.title('Top 10 Countries with Most Content')
plt.xlabel('Number of Titles')
plt.show()
```



In [44]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
dataset['rating']=le.fit_transform(dataset['rating'])
```

In [45]:
```
plt.figure(figsize=(12, 6))
sns.boxplot(x='rating', y='duration_mintues', data=df)
plt.title('Rating vs Duration')
plt.xlabel('Rating')
plt.ylabel('Duration (minutes)')
plt.show()
```



In [ ]: