In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt
df=pd.read_csv("dataset - netflix1.csv")
df
```

Out[1]:

	show_id	type	title	director	country	date_added	release_year	rating	dura
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	Se
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	Se
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125
...	...	...	...	...	...	...	...	...	
8785	s8797	TV Show	Yunus Emre	Not Given	Turkey	1/17/2017	2016	TV-PG	Sea
8786	s8798	TV Show	Zak Storm	Not Given	United States	9/13/2018	2016	TV-Y7	Sea
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	Pakistan	12/15/2016	2012	TV-PG	Se
8788	s8784	TV Show	Yoko	Not Given	Pakistan	6/23/2018	2016	TV-Y	Se
8789	s8786	TV Show	YOM	Not Given	Pakistan	6/7/2018	2016	TV-Y7	Se

8790 rows × 10 columns

In [2]:
```
df.head()
```

Out[2]:

	show_id	type	title	director	country	date_added	release_year	rating	duration
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min

In [3]:
```
df.isnull().sum()
```

Out[3]:
```
show_id         0
type            0
title           0
director        0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
dtype: int64
```

In [4]:
```
df.describe()
```

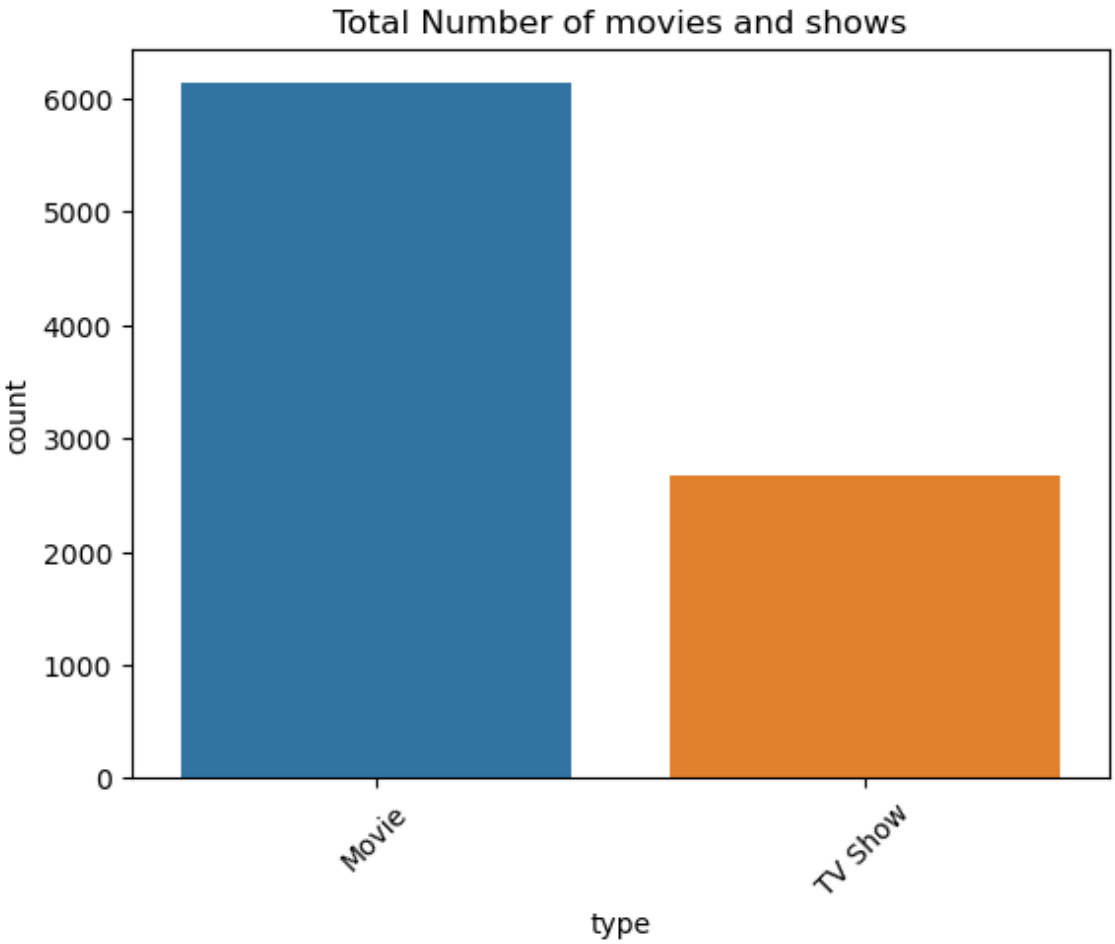
Out[4]:

	release_year
count	8790.000000
mean	2014.183163
std	8.825466
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

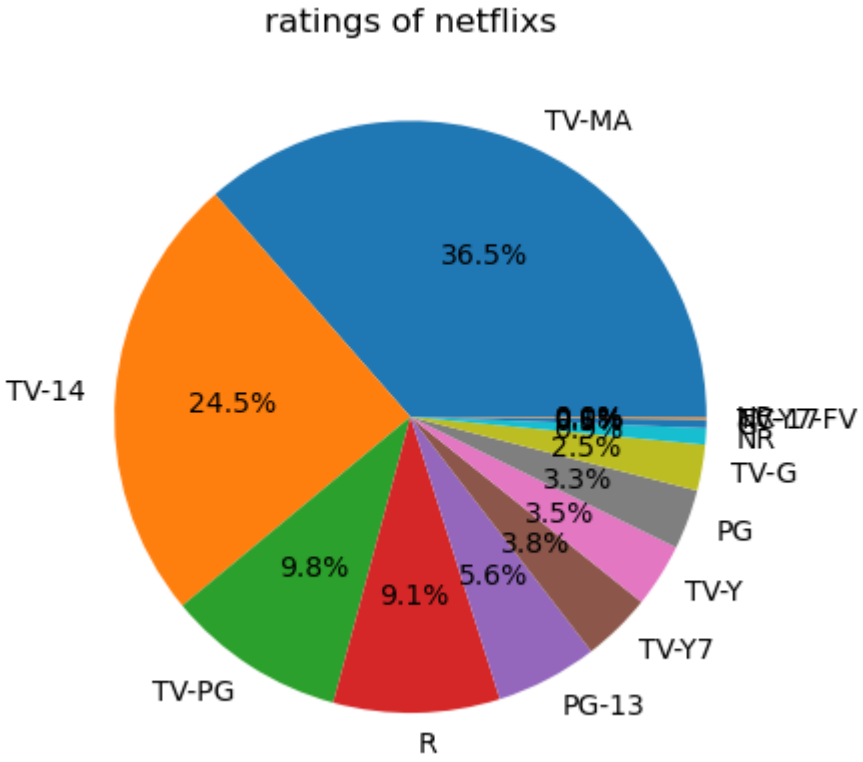
In [5]:
```
df.shape
```

Out[5]:
```
(8790, 10)
```

In [6]:
```
sn.countplot(x='type',data=df)
plt.title('Total Number of movies and shows')
plt.xticks(rotation=45)
plt.show()
```



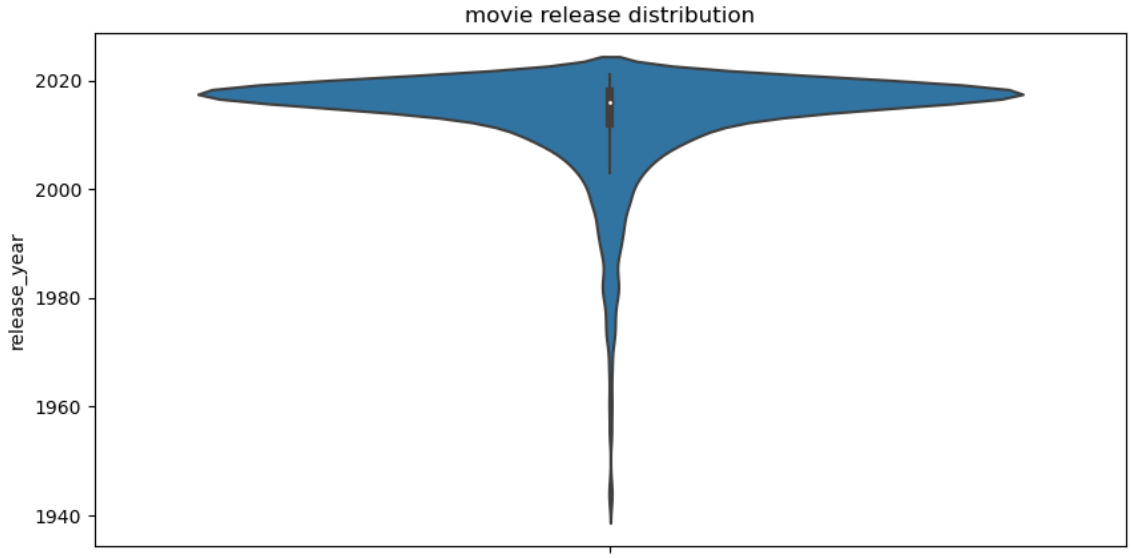
In [7]:
```
rat_counts=df[ 'rating'].value_counts()
plt.pie(rat_counts, labels=rat_counts.index, autopct='%1.1f%%')
plt.title('ratings of netflixs')
plt.show()
```



In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8790 non-null   object
 1   type          8790 non-null   object
 2   title         8790 non-null   object
 3   director      8790 non-null   object
 4   country       8790 non-null   object
 5   date_added    8790 non-null   object
 6   release_year  8790 non-null   int64 
 7   rating        8790 non-null   object
 8   duration      8790 non-null   object
 9   listed_in     8790 non-null   object
dtypes: int64(1), object(9)
memory usage: 686.8+ KB
```

In [9]:
```
df['relaese_year']=df[ 'release_year'].astype(int)
plt.figure(figsize=(10,5))
sn.violinplot(y= 'release_year', data=df[df['type'] == 'Movie'])
plt.title('movie release distribution')
plt.ylabel('release_year')
plt.show()
```



In [ ]:
```

```