aicte-project-2-1

November 20, 2023

[1]:
```
#PHASE 2 TASK1
```

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```
data=pd.read_csv(r"C:\Users\Prava\OneDrive\Desktop\dataset - netflix1.csv")
data
```

[3]:

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	s3	TV Show	Ganglands	Julien Leclercq	
2	s6	TV Show	Midnight Mass	Mike Flanagan	
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	
4	s8	Movie	Sankofa	Haile Gerima	
…	…	…	…	…	
8785	s8797	TV Show	Yunus Emre	Not Given	
8786	s8798	TV Show	Zak Storm	Not Given	
8787	s8801	TV Show	Zindagi Gulzar Hai	Not Given	
8788	s8784	TV Show	Yoko	Not Given	
8789	s8786	TV Show	YOM	Not Given	

	country	date_added	release_year	rating	duration	\
0	United States	9/25/2021	2020	PG-13	90 min	
1	France	9/24/2021	2021	TV-MA	1 Season	
2	United States	9/24/2021	2021	TV-MA	1 Season	
3	Brazil	9/22/2021	2021	TV-PG	91 min	
4	United States	9/24/2021	1993	TV-MA	125 min	
…	…	…	…	…	…	
8785	Turkey	1/17/2017	2016	TV-PG	2 Seasons	
8786	United States	9/13/2018	2016	TV-Y7	3 Seasons	
8787	Pakistan	12/15/2016	2012	TV-PG	1 Season	
8788	Pakistan	6/23/2018	2016	TV-Y	1 Season	
8789	Pakistan	6/7/2018	2016	TV-Y7	1 Season	

	listed_in
0	Documentaries

```
1     Crime TV Shows, International TV Shows, TV Act…
2                   TV Dramas, TV Horror, TV Mysteries
3                   Children & Family Movies, Comedies
4      Dramas, Independent Movies, International Movies
…                                                    …
8785                 International TV Shows, TV Dramas
8786                                         Kids' TV
8787  International TV Shows, Romantic TV Shows, TV …
8788                                         Kids' TV
8789                                         Kids' TV

[8790 rows x 10 columns]
```

[4]:
```
#READ THE CLEANED DATASET
```

[5]:
```
data1=pd.read_csv(r"C:\Users\Prava\OneDrive\Desktop\dataset - netflix1.csv")
data1
```

[5]:
```
     show_id     type                             title         director  \
0         s1    Movie              Dick Johnson Is Dead  Kirsten Johnson
1         s3  TV Show                         Ganglands  Julien Leclercq
2         s6  TV Show                     Midnight Mass    Mike Flanagan
3        s14    Movie  Confessions of an Invisible Girl    Bruno Garotti
4         s8    Movie                           Sankofa     Haile Gerima
…       …       …                                …               …
8785   s8797  TV Show                       Yunus Emre        Not Given
8786   s8798  TV Show                        Zak Storm        Not Given
8787   s8801  TV Show                Zindagi Gulzar Hai        Not Given
8788   s8784  TV Show                             Yoko        Not Given
8789   s8786  TV Show                              YOM        Not Given

            country  date_added  release_year rating   duration  \
0     United States   9/25/2021          2020  PG-13     90 min
1            France   9/24/2021          2021  TV-MA   1 Season
2     United States   9/24/2021          2021  TV-MA   1 Season
3            Brazil   9/22/2021          2021  TV-PG     91 min
4     United States   9/24/2021          1993  TV-MA    125 min
…               …         …             …     …        …
8785         Turkey   1/17/2017          2016  TV-PG  2 Seasons
8786  United States   9/13/2018          2016  TV-Y7  3 Seasons
8787       Pakistan  12/15/2016          2012  TV-PG   1 Season
8788       Pakistan   6/23/2018          2016   TV-Y   1 Season
8789       Pakistan    6/7/2018          2016  TV-Y7   1 Season

                                         listed_in
0                                    Documentaries
1     Crime TV Shows, International TV Shows, TV Act…
```

```
2                    TV Dramas, TV Horror, TV Mysteries
3                    Children & Family Movies, Comedies
4       Dramas, Independent Movies, International Movies
…                                                     …
8785                  International TV Shows, TV Dramas
8786                                          Kids' TV
8787  International TV Shows, Romantic TV Shows, TV …
8788                                          Kids' TV
8789                                          Kids' TV

[8790 rows x 10 columns]
```

#Exploratory Data Analysis

[6]:
```
data1.head()
```

[6]:
```
  show_id     type                              title         director  \
0      s1    Movie              Dick Johnson Is Dead  Kirsten Johnson
1      s3  TV Show                         Ganglands  Julien Leclercq
2      s6  TV Show                     Midnight Mass    Mike Flanagan
3     s14    Movie  Confessions of an Invisible Girl    Bruno Garotti
4      s8    Movie                           Sankofa     Haile Gerima

         country date_added  release_year rating  duration  \
0  United States  9/25/2021          2020  PG-13    90 min
1         France  9/24/2021          2021  TV-MA  1 Season
2  United States  9/24/2021          2021  TV-MA  1 Season
3         Brazil  9/22/2021          2021  TV-PG    91 min
4  United States  9/24/2021          1993  TV-MA   125 min

                                        listed_in
0                                   Documentaries
1  Crime TV Shows, International TV Shows, TV Act…
2                TV Dramas, TV Horror, TV Mysteries
3                Children & Family Movies, Comedies
4   Dramas, Independent Movies, International Movies
```

[7]:
```
data1.tail()
```

[7]:
```
     show_id     type               title   director        country  \
8785   s8797  TV Show          Yunus Emre  Not Given         Turkey
8786   s8798  TV Show           Zak Storm  Not Given  United States
8787   s8801  TV Show  Zindagi Gulzar Hai  Not Given       Pakistan
8788   s8784  TV Show                Yoko  Not Given       Pakistan
8789   s8786  TV Show                 YOM  Not Given       Pakistan

      date_added  release_year rating   duration  \
```

```
8785   1/17/2017          2016  TV-PG  2 Seasons
8786   9/13/2018          2016  TV-Y7  3 Seasons
8787  12/15/2016          2012  TV-PG   1 Season
8788   6/23/2018          2016   TV-Y   1 Season
8789    6/7/2018          2016  TV-Y7   1 Season

                                              listed_in
8785                  International TV Shows, TV Dramas
8786                                           Kids' TV
8787  International TV Shows, Romantic TV Shows, TV …
8788                                           Kids' TV
8789                                           Kids' TV
```

[8]:
```
data1.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
```

[9]:
```
data1.describe()
```

[9]:
```
       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
```

[10]:
```
data1.type.value_counts()
```

[10]:
```
Movie      6126
TV Show    2664
Name: type, dtype: int64
```

[11]:
```
data1.title.value_counts()
```

[11]:
```
9-Feb                   2
15-Aug                  2
22-Jul                  2
Dick Johnson Is Dead    1
SGT. Will Gardner       1
                       ..
Mercy Black             1
The Trap                1
Pinky Memsaab           1
Love 020                1
YOM                     1
Name: title, Length: 8787, dtype: int64
```

[12]:
```
data1.country.value_counts()
```

[12]:
```
United States     3240
India             1057
United Kingdom     638
Pakistan           421
Not Given          287
                   …
Iran                  1
West Germany          1
Greece                1
Zimbabwe              1
Soviet Union          1
Name: country, Length: 86, dtype: int64
```

[13]:
```
data1.date_added.value_counts()
```

[13]:
```
1/1/2020      110
11/1/2019      91
3/1/2018       75
12/31/2019     74
10/1/2018      71
               …
6/26/2015       1
6/23/2015       1
6/1/2015        1
5/29/2015       1
4/1/2014        1
```

```
Name: date_added, Length: 1713, dtype: int64
```

[14]:
```
data1.release_year.value_counts()
```

[14]:
```
2018    1146
2017    1030
2019    1030
2020     953
2016     901
        …
1966       1
1959       1
1925       1
1947       1
1961       1
Name: release_year, Length: 74, dtype: int64
```

[15]:
```
data1.rating.value_counts()
```

[15]:
```
TV-MA       3205
TV-14       2157
TV-PG        861
R            799
PG-13        490
TV-Y7        333
TV-Y         306
PG           287
TV-G         220
NR            79
G             41
TV-Y7-FV       6
NC-17          3
UR             3
Name: rating, dtype: int64
```

[16]:
```
data1.duration.value_counts()
```

[16]:
```
1 Season      1791
2 Seasons      421
3 Seasons      198
90 min         152
97 min         146
              …
5 min            1
16 min           1
186 min          1
193 min          1
```

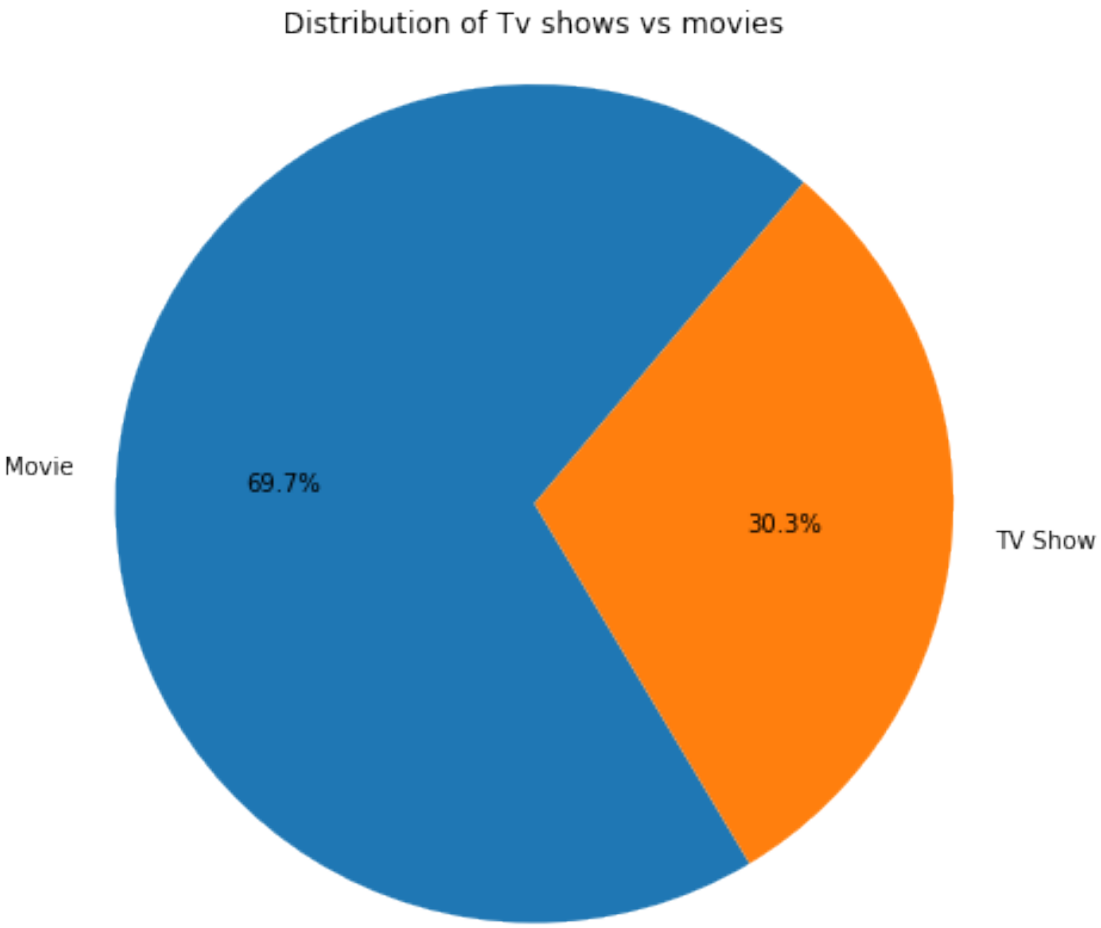
```
11 Seasons       1
Name: duration, Length: 220, dtype: int64
```

[17]:
```
data1.listed_in.value_counts()
```

[17]:
```
Dramas, International Movies                                362
Documentaries                                               359
Stand-Up Comedy                                             334
Comedies, Dramas, International Movies                      274
Dramas, Independent Movies, International Movies            252
                                                           ...
Anime Features                                                1
Action & Adventure, Horror Movies, Independent Movies         1
Action & Adventure, Classic Movies, International Movies      1
Cult Movies, Independent Movies, Thrillers                    1
Classic & Cult TV, Crime TV Shows, TV Dramas                  1
Name: listed_in, Length: 513, dtype: int64
```

[18]:
```
#Analysing the data using matplotlib
```

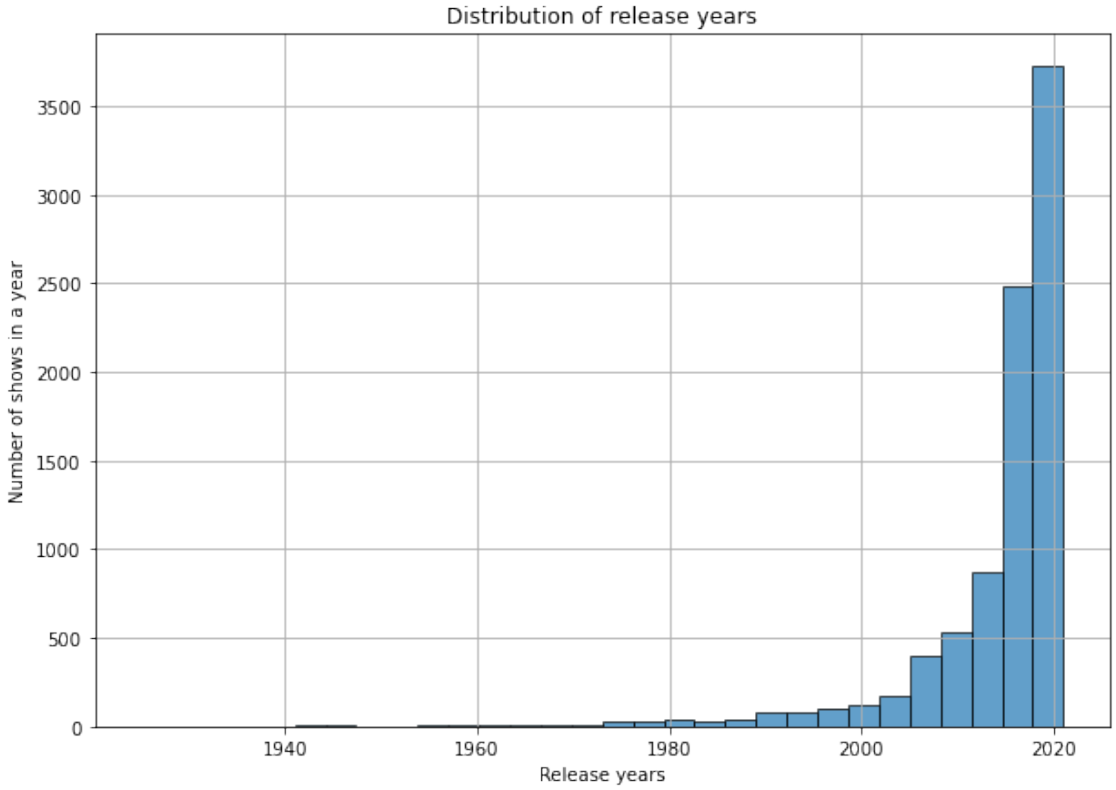
[19]:
```
counts=data1.type.value_counts()
labels=counts.index
sizes=counts.values
plt.figure(figsize=(7,7))
plt.pie(sizes,labels=labels,autopct='%1.1f%%',startangle=50)
plt.axis('equal')
plt.title("Distribution of Tv shows vs movies")
plt.show()
```



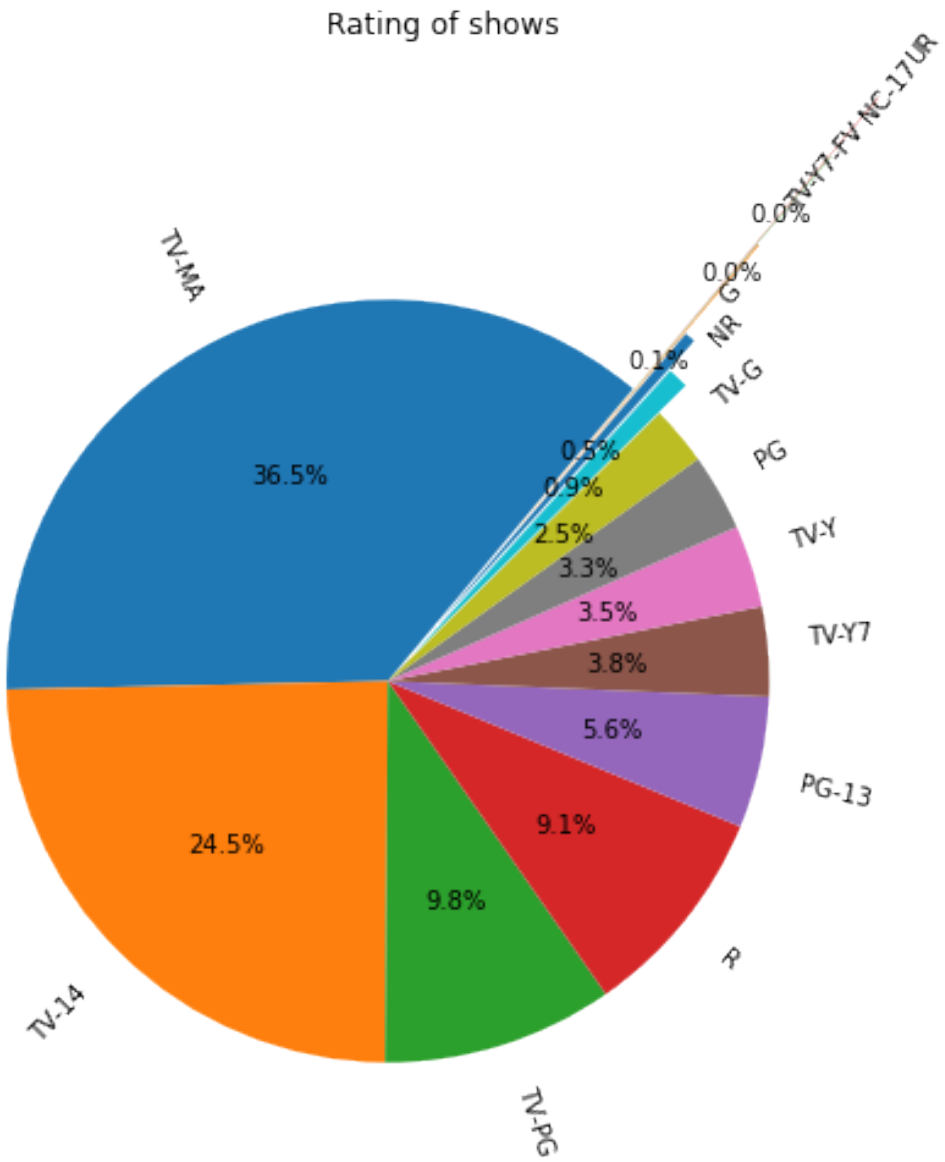
[20]:
```
# Interpretation:Netflix has a large range of movies compared to tv shows
```

[21]:
```
#histogram to analyze the number ofshows released in each year
```

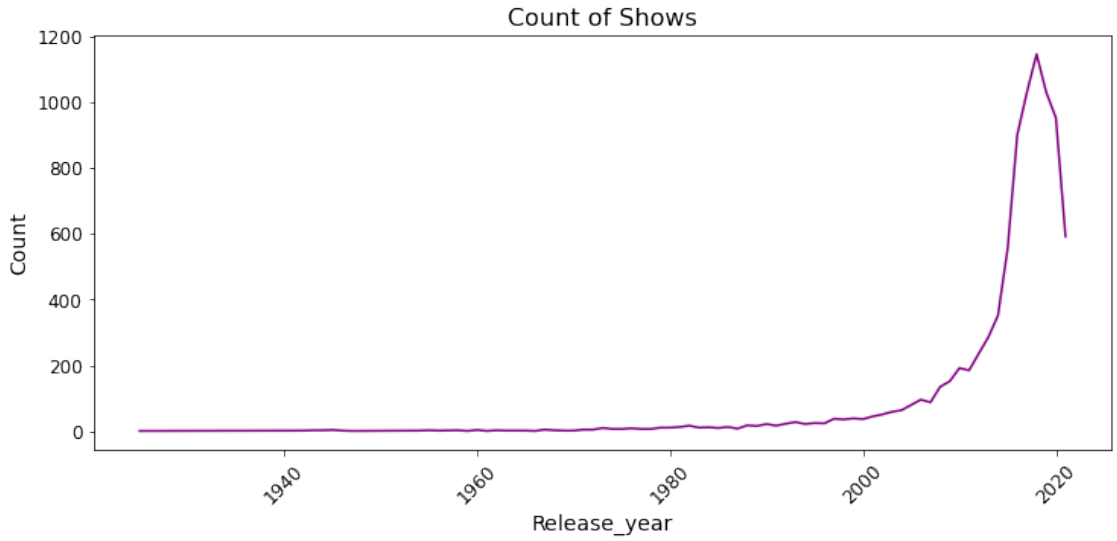
[22]:
```
plt.figure(figsize=(10,7))
plt.hist(data1['release_year'],bins=30,edgecolor='black',alpha=0.7)
plt.ylabel("Number of shows in a year")
plt.xlabel("Release years")
plt.title("Distribution of release years")
plt.grid(True)
plt.show()
```



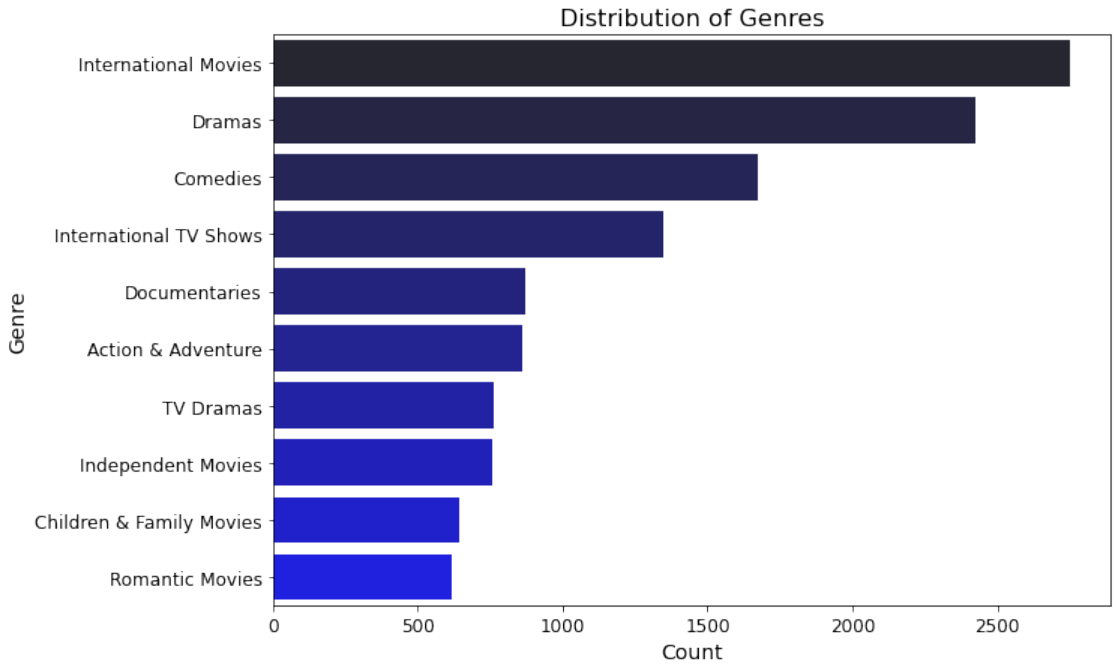
[24]:
```
#Pie chart to analyze the rating of show
counts1=data1.rating.value_counts()
labels=counts1.index
sizes=counts1.values
plt.figure(figsize=(8, 8))
plt.pie(sizes, labels=labels, autopct='%1.1f%%',explode=(0,0,0,0,0,0,0,0,0,0.
↪1,0.2,0.5,0.8,1),startangle=50,rotatelabels=21)
plt.axis('equal')
plt.title("Rating of shows")
plt.show()
```



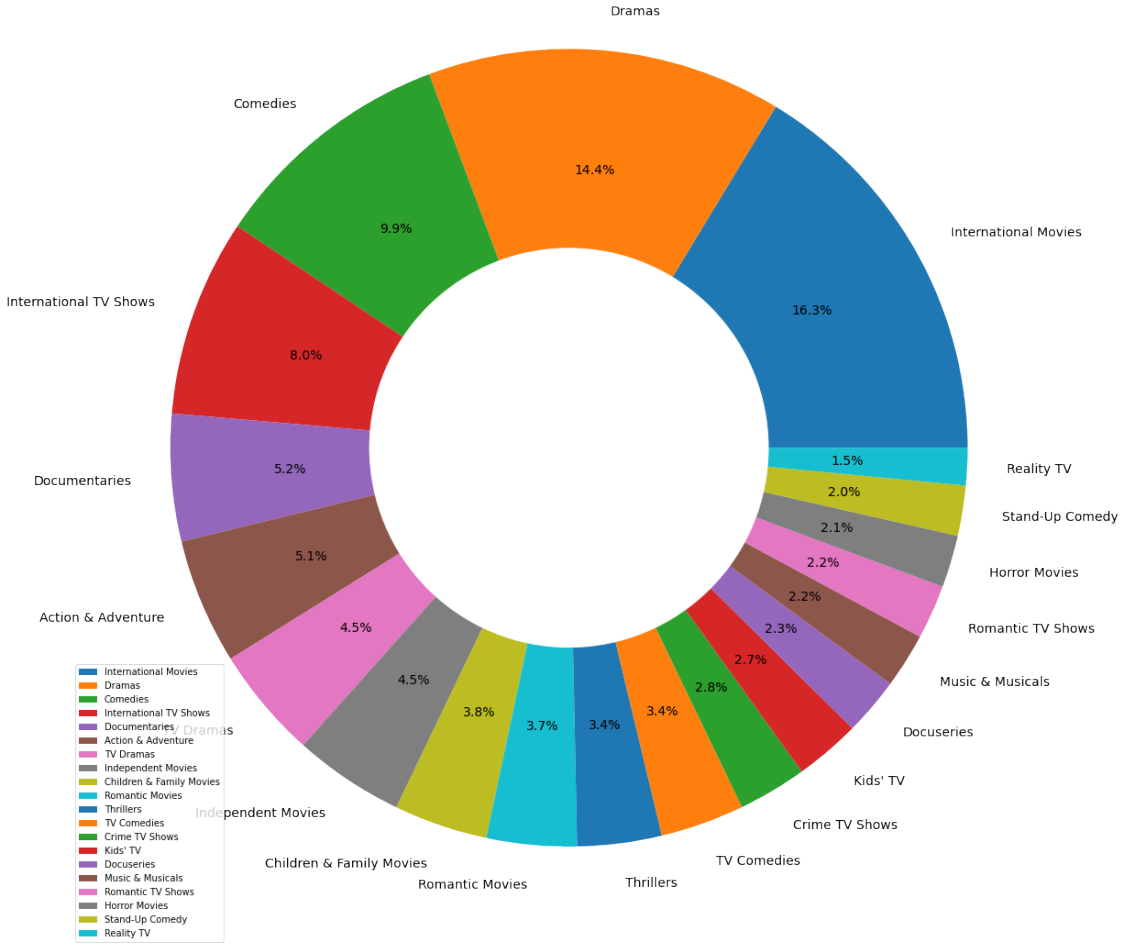
[39]:
```
plt.figure(figsize=(12,5))
shows_per_year = data1.groupby('release_year').size().reset_index(name='count')
sns.lineplot(data=shows_per_year, x='release_year', y='count', color='purple')
plt.xlabel('Release_year', fontsize=14)
plt.ylabel('Count', fontsize=14)
plt.title('Count of Shows', fontsize=16)
plt.xticks(rotation=45, fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```



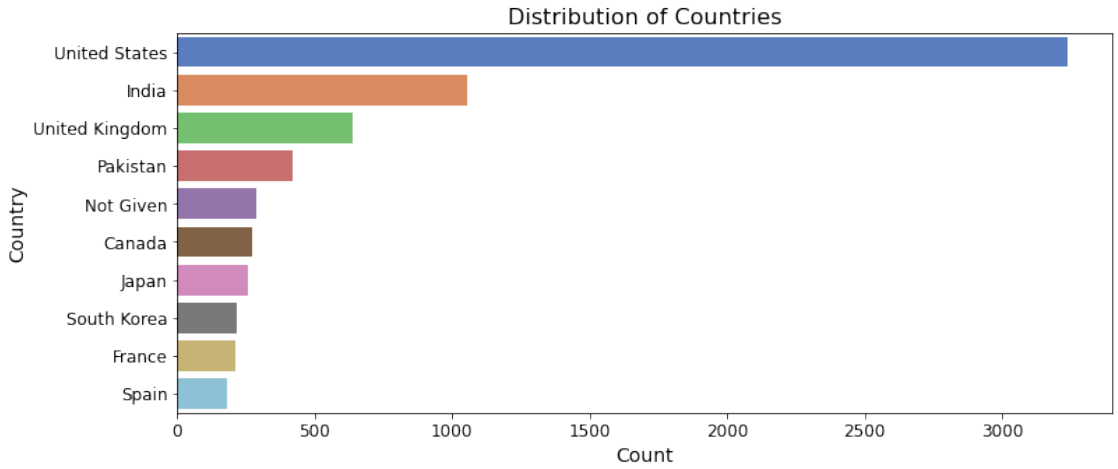
[32]:
```
plt.figure(figsize=(10, 7))
genres = data1['listed_in'].str.split(', ').explode().str.strip()
top_genres = genres.value_counts().head (10)
sns.barplot(x=top_genres.values, y=top_genres.index, palette='dark:blue')
plt.xlabel('Count', fontsize=14)
plt.ylabel('Genre', fontsize=14)
plt.title('Distribution of Genres', fontsize=16)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```



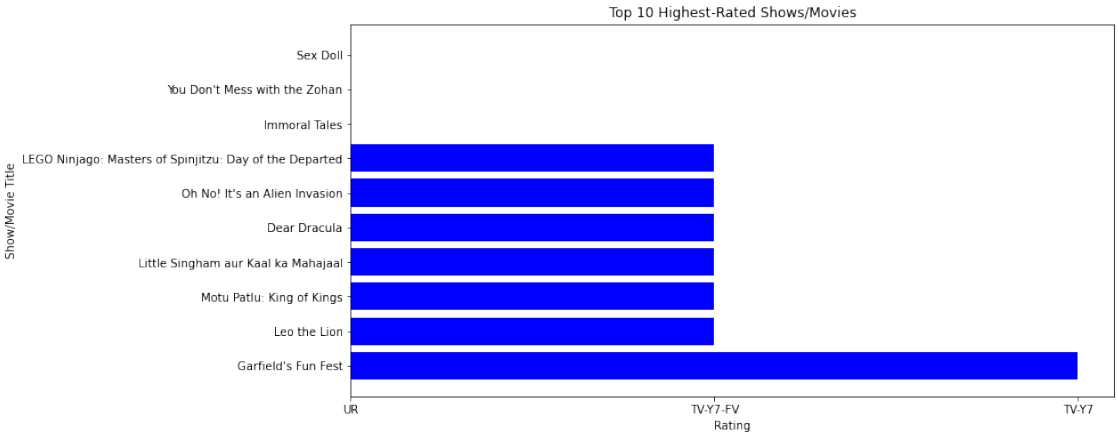
[29]:
```
genre_counts_top20 = sorted(genre_counts.items(), key=lambda item:item[1],␣
↪reverse = True)[:20]
genre_counts_top20 = dict(genre_counts_top20)

labels = genre_counts_top20.keys()
sizes = genre_counts_top20.values()
inner_circle_radius = 0.5

plt.figure(figsize=(30, 20))
plt.pie(sizes,autopct='%1.1f%%',labels=labels,pctdistance=0.
↪7,textprops={'fontsize': 14})
plt.gca().set_aspect('equal')

inner_circle = plt.Circle((0, 0), inner_circle_radius, color='white')
plt.gca().add_artist(inner_circle)
plt.legend()
plt.title('Categories',fontsize=23)
plt.show()
```



[41]:
```
plt.figure(figsize=(12, 5))
top_countries =data1['country'].value_counts().head (10)
sns.barplot(x=top_countries.values, y=top_countries.index, palette='muted')
plt.xlabel('Count', fontsize=14)
plt.ylabel('Country', fontsize=14)
plt.title('Distribution of Countries', fontsize=16)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```



[44]:
```
sorted_data1 = data1.sort_values(by='rating', ascending=False)
top_n = 10
top_rated = sorted_data1.head(top_n)

plt.figure(figsize=(12, 6))
plt.barh(top_rated['title'], top_rated['rating'], color='blue')
plt.xlabel('Rating')
plt.ylabel('Show/Movie Title')
plt.title('Top {} Highest-Rated Shows/Movies'.format(top_n))
plt.gca().invert_yaxis()
plt.show()
```



[ ]: