PRACTICAL 4

Data science and Visualization

NAME: AAYUSHI DIGHE

PRN: 72017865H

CLASS: TE ENTC A

In this practical we will perform Data Visualization.

In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv(r"C:\Users\Hp\Downloads\netflix_titles.csv")
df.head(8807)
```

Out[2]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA

	show_id	type	title	director	cast	country	date_added	release_year	rating
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA
...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14

8807 rows × 12 columns

In [3]:
```
df.shape
```

Out[3]: (8807, 12)

In [4]:
```
categories=df['listed_in']
```

In [5]:
```
total_child=sum(df['listed_in'].str.contains('Child'))
```

In [6]:
```
total_child
```

Out[6]: 641

In [7]:
```
Standup_Comedies=sum(df['listed_in'].str.contains('Stand'))
```

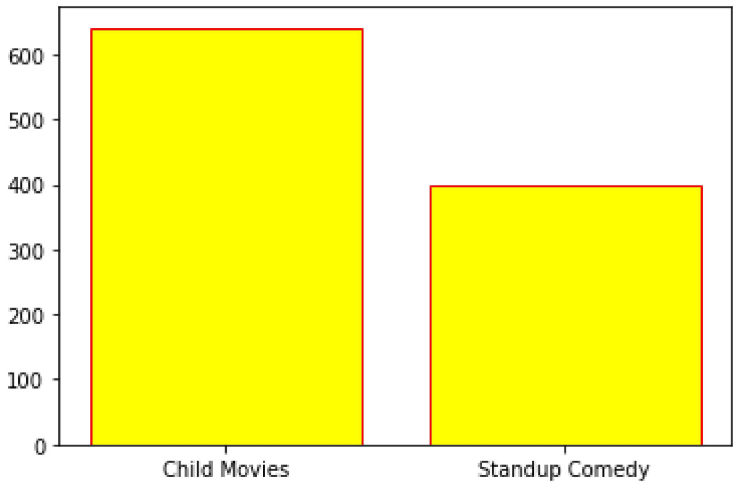
In [8]:
```
Standup_Comedies
```

Out[8]: 399

We determined the number of child movies/shows and standup comedies.We will visualize this number using plot.

In [9]:
```
import matplotlib.pyplot as plt
```

In [37]:
```
plt.bar(['Child Movies','Standup Comedy'],
       [total_child, Standup_Comedies],color='yellow',edgecolor='red')
plt.show()
```



In [11]:
```
set(df['type'])
```

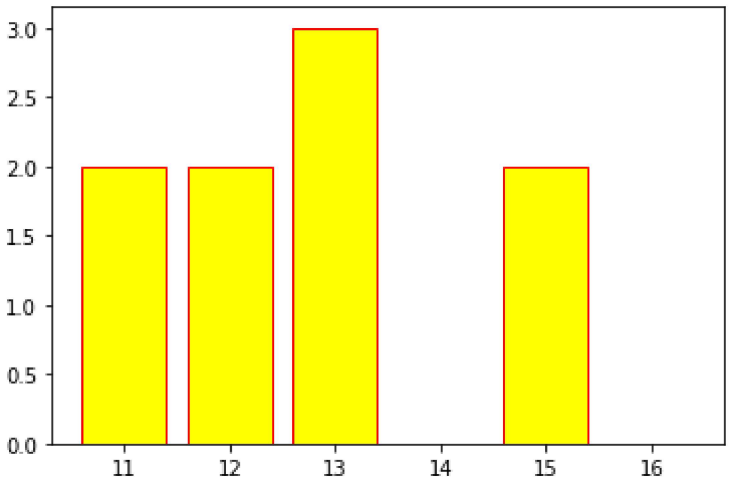
Out[11]: {'Movie', 'TV Show'}

In [12]:
```
tv_shows = df[df['type'] == 'TV Show'] #Boolean Filtering
```

In [13]:
```
seasons13 = tv_shows [tv_shows [ 'duration'] == '13 Seasons']

seasons15 = tv_shows [tv_shows['duration'] == '15 Seasons']

seasons16= tv_shows [tv_shows['duration'] == '16 Seasons']

seasons12 = tv_shows [tv_shows['duration'] == '12 Seasons']

seasons11= tv_shows [tv_shows['duration'] == '11 Seasons']
```

In [26]:
```
plt.bar ([11, 12, 13, 15, 16],
[len(seasons11), len(seasons12), len(seasons13), len (seasons15), len(seasons16)],

color='yellow',edgecolor='red')
```

Out[26]: <BarContainer object of 5 artists>

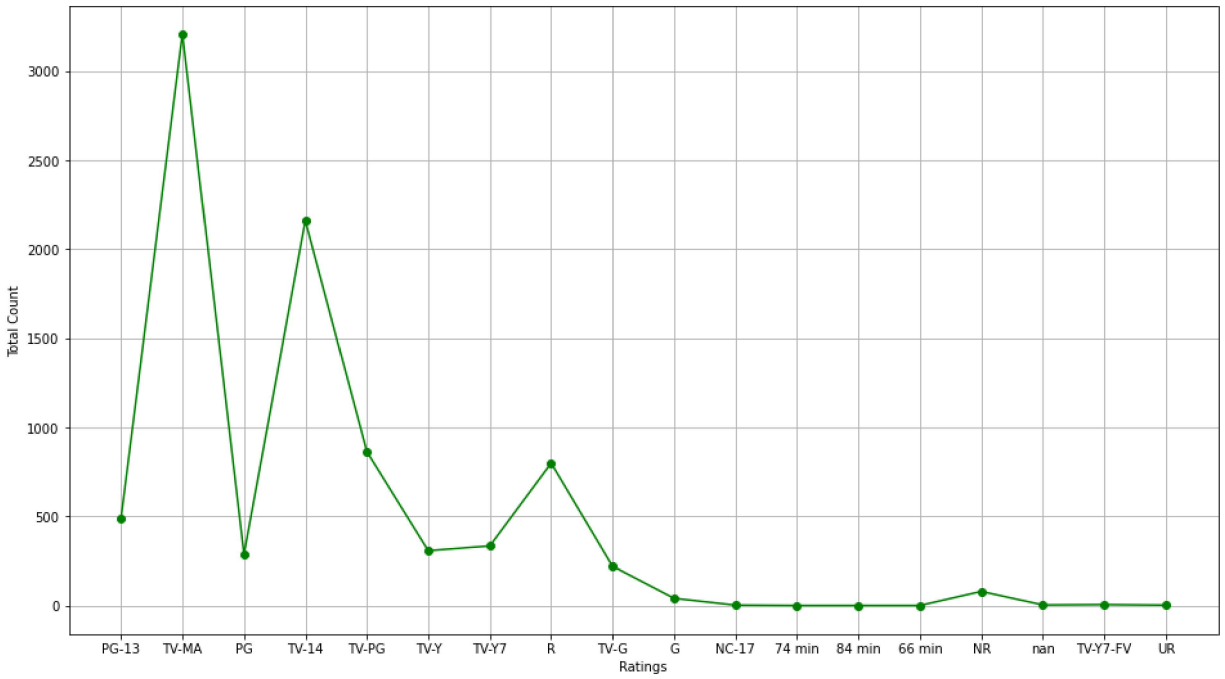


In [15]:
```
from collections import Counter
ratings = Counter(df['rating'])
```

In [16]:
```
ratings
```

Out[16]:
```
Counter({'PG-13': 490,
         'TV-MA': 3207,
         'PG': 287,
         'TV-14': 2160,
         'TV-PG': 863,
         'TV-Y': 307,
         'TV-Y7': 334,
         'R': 799,
         'TV-G': 220,
         'G': 41,
         'NC-17': 3,
         '74 min': 1,
         '84 min': 1,
         '66 min': 1,
         'NR': 80,
         nan: 4,
         'TV-Y7-FV': 6,
         'UR': 3})
```

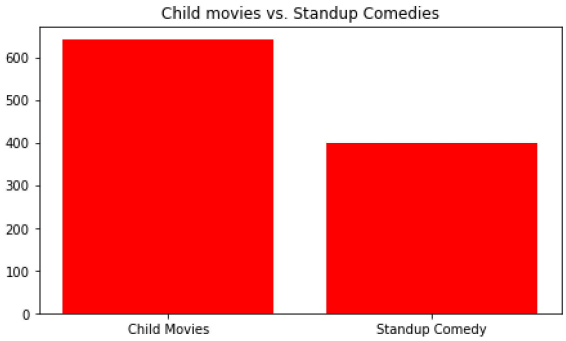
In [35]:
```
plt. figure(figsize=(16,9))
plt.plot(ratings.keys(), ratings.values(), color = 'green', marker='o')
plt.xlabel('Ratings'); plt.ylabel('Total Count')
plt.grid()
```

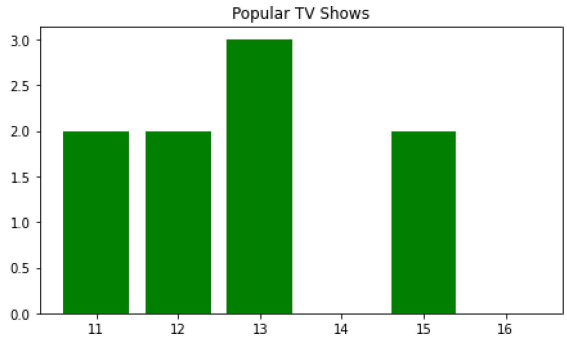


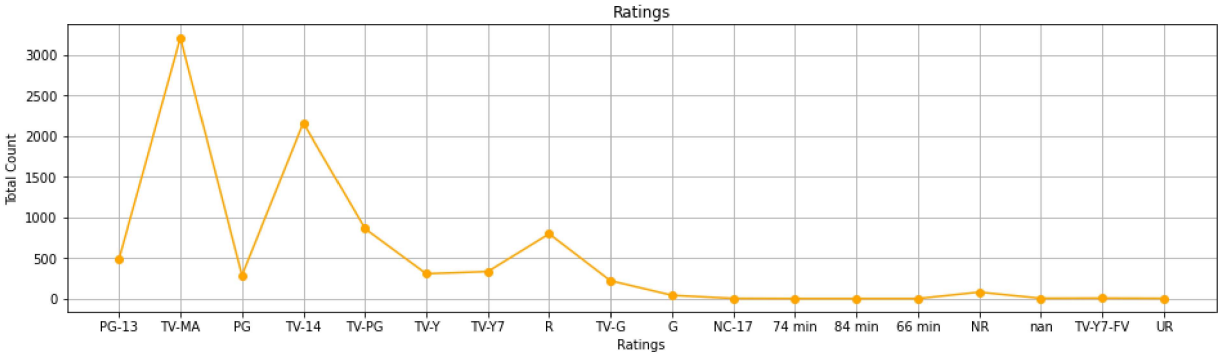
If we wish to plot all these plots in the same plot we can use subplot.

In [40]:

```
plt. figure (figsize=(16,9))

#plot1
plt.subplot (2,2,1)
plt.title ("Child movies vs. Standup Comedies")
plt.bar(['Child Movies', 'Standup Comedy'], [total_child, Standup_Comedies],color="r

#plot2
plt. subplot (2,2,2)
plt.title('Popular TV Shows')
plt.bar([11, 12, 13, 15, 16],
[len (seasons11), len (seasons12), len(seasons13),
len (seasons15), len (seasons16)],
color='green')

#plot3
plt.subplot (2,1,2)
plt.title('Ratings')
plt.plot(ratings.keys (), ratings.values(), color='orange', marker='o')
plt.xlabel('Ratings'); plt.ylabel('Total Count')
plt.grid()
```


Child movies vs. Standup Comedies
600
500
400
300
200
100
0
Child Movies
Standup Comedy


Popular TV Shows
3.0
2.5
2.0
1.5
1.0
0.5
0.0
11
12
13
14
15
16


Ratings
Total Count
3000
2500
2000
1500
1000
500
0
PG-13
TV-MA
PG
TV-14
TV-PG
TV-Y
TV-Y7
R
TV-G
G
NC-17
74 min
84 min
66 min
NR
nan
TV-Y7-FV
UR
Ratings

In [ ]: