

In [69]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
pd.options.display.float_format = '{:.2f}'.format
```

In [2]:

```
df = pd.read_csv('../data/zippedData/rt.movie_info.tsv.gz', delimiter='\t')
df.head()
```

Out[2]:

	id	synopsis	rating	genre	director	writer	theater_date	dvd_dat
0	1	This gritty, fast-paced, and innovative police...	R	Action and Adventure Classics Drama	William Friedkin	Ernest Tidyman	Oct 9, 1971	Sep 21, 200
1	3	New York City, not-too-distant-future: Eric Pa...	R	Drama Science Fiction and Fantasy	David Cronenberg	David Cronenberg Don DeLillo	Aug 17, 2012	Jan 1, 201
2	5	Illeana Douglas delivers a superb performance	R	Drama Musical and Performing Arts	Allison Anders	Allison Anders	Sep 13, 1996	Apr 1, 200
		...						
3	6	Michael Douglas runs afoul of a treacherous su...	R	Drama Mystery and Suspense	Barry Levinson	Paul Attanasio Michael Crichton	Dec 9, 1994	Aug 2, 199
4	7	NaN	NR	Drama Romance	Rodney Bennett	Giles Cooper	NaN	NaN

In [3]:

```
df.info()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1560 entries, 0 to 1559
Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	id	1560	non-null int64
1	synopsis	1498	non-null object
2	rating	1557	non-null object
3	genre	1552	non-null object
4	director	1361	non-null object
5	writer	1111	non-null object
6	theater_date	1201	non-null object
7	dvd_date	1201	non-null object
8	currency	340	non-null object
9	box_office	340	non-null object
10	runtime	1530	non-null object
11	studio	494	non-null object

```
dtypes: int64(1), object(11)
memory usage: 146.4+ KB
```

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

```
Out[4]: count      340
unique     336
top    200,000
freq       2
Name: box_office, dtype: object
```

```
In [5]: box_office_with_non_value = df.loc[df['box_office'].notnull()]
box_office_with_non_value
```

		id	synopsis	rating	genre	director	writer	theater_date
1	3	New York City, not- too-distant- future: Eric Pa...	R	Drama Science Fiction and Fantasy	David Cronenberg	David Cronenberg Don DeLillo	Aug 17, 2012	
6	10	Some cast and crew from NBC's highly acclaimed...	PG-13	Comedy	Jake Kasdan	Mike White	Jan 11, 2002	
7	13	Stewart Kane, an Irishman living in the Austra...	R	Drama	Ray Lawrence	Raymond Carver Beatrix Christian	Apr 27, 2006	
8	14	"Love Ranch" is a bittersweet love story that ...	R	Drama	Taylor Hackford	Mark Jacobson	Jun 30, 2010	
15	22	Two-time Academy Award Winner Kevin Spacey giv...	R	Comedy Drama Mystery and Suspense	George Hickenlooper	Norman Snider	Dec 17, 2010	
...	
1541	1980	A band of renegades on the run in outer space ...	PG-13	Action and Adventure Science Fiction and Fantasy	Joss Whedon	Joss Whedon	Sep 30, 2005	
1542	1981	Money, Fame and the Knowledge of English. In I...	NR	Comedy Drama	Gauri Shinde	Gauri Shinde	Oct 5, 2012	

	id	synopsis	rating	genre	director	writer	theater_date
1545	1985	A woman who joins the undead against her will ...	R	Horror Mystery and Suspense	Sebastian Gutierrez	Sebastian Gutierrez	Jun 1, 2007
1546	1986	Aki Kaurismaki's The Man Without a Past opens ...	PG	Art House and International Comedy Drama	NaN	NaN	Aug 30, 2002
1555	1996	Forget terrorists or hijackers -- there's a ha...	R	Action and Adventure Horror Mystery and Suspense	NaN	NaN	Aug 18, 2006

340 rows × 12 columns



```
In [6]: bownv = box_office_with_non_value.sort_values( by=['box_office'], ascending=False)
bownv
```

	id	synopsis	rating	genre	director	writer	theater_da
16	23	A fictional film set in the alluring world of ...	R	Drama	NaN	NaN	Dec 20, 20
1510	1944	In the 3D action-comedy The Green Hornet, Brit...	PG-13	Action and Adventure Comedy Science Fiction an...	Michel Gondry	Evan Goldberg Seth Rogen	Jan 14, 20
207	274	Natalie Portman stars in this western concerni...	R	Action and Adventure Drama Western	NaN	NaN	Jan 29, 20
1469	1890	A misplaced sausage and his savory friends emb...	R	Animation Comedy	Conrad Vernon Greg Tiernan	Kyle Hunter Ariel Shaffir Seth Rogen Evan Gold...	Aug 12, 20
1323	1711	Directed by Garry Marshall, The Princess Diari...	G	Comedy Kids and Family	Garry Marshall	Shonda Rhimes	Aug 11, 20
...
1213	1564	Real-life sweethearts Michael Cera and Charl...	PG-13	Comedy	Nicholas Jasenovec	Nicholas Jasenovec Charlyne Yi	Aug 14, 20

	id	synopsis	rating	genre	director	writer	theater_date
1087	1398	A heartfelt tale of inspiration, hope and rede...	PG	Drama Kids and Family Faith and Spirituality	David Nixon Patrick Doughtie	NaN	Apr 9, 20
809	1046	Writer/director Rodrigo Garca (Nine Lives) tea...	R	Drama	NaN	NaN	May 7, 20
15	22	Two-time Academy Award Winner Kevin Spacey giv...	R	Comedy Drama Mystery and Suspense	George Hickenlooper	Norman Snider	Dec 17, 20
86	108	A picture-perfect family moves into an upscale...	R	Comedy Drama Kids and Family	Derrick Borte	Derrick Borte	Apr 16, 20

340 rows × 12 columns



```
In [7]: bownv['box_office'] = bownv['box_office'].str.replace(',', '')
```

```
In [8]: bownv['box_office'] = bownv['box_office'].astype(int)
bownv['box_office']
```

```
Out[8]: 16      99165609
1510    98000000
207     976847
1469    97661826
1323    95111012
...
1213    1149350
1087    1121000
809     1110286
15      1039869
86     1000000
Name: box_office, Length: 340, dtype: int64
```

```
In [9]: bownv.sort_values('box_office', ascending=False, inplace=True)
```

	id	synopsis	rating	genre	director	writer	theater_date	dvd
827	1071	This depiction of the last 12 hours in the life...	R	Drama	Mel Gibson	Mel Gibson Benedict Fitzgerald	Feb 25, 2004	A
877	1136	The Hobbit: An Unexpected Journey follows titl...	PG-13	Action and Adventure Science Fiction and Fantasy	Peter Jackson	Peter Jackson Philippa Boyens Guillermo del To...	Dec 14, 2012	M

	id	synopsis	rating	genre	director	writer	theater_date	dvd
1188	1525	In Skyfall, Bond's loyalty to M is tested as h...	PG-13	Action and Adventure Mystery and Suspense	Sam Mendes	John Logan Robert Wade Neal Purvis	Nov 9, 2012	M
690	883	Now that Greg Focker is "in" with his soon-to...	PG-13	Comedy	Jay Roach	Jim Herzfeld John Hamburg	Dec 22, 2004	A
1117	1442	Youthful wizard Harry Potter returns to the sc...	PG	Kids and Family Science Fiction and Fantasy	Chris Columbus	Steve Kloves	Nov 15, 2002	A

1269	1639	In this darkly erotic drama from cult filmmaker...	NR	Art House and International Drama	Jeanne Labrune	Jeanne Labrune Richard Debusine	Aug 26, 2011	J
691	884	Freebie (James Caan) and the Bean (Alan Arkin)...	R	Action and Adventure Comedy Drama	Richard Rush	Robert Kaufman	Dec 25, 1974	J
697	895	After a sexual assault by a random stranger, S...	NR	Drama Horror Mystery and Suspense	Eric England	Eric England	Nov 22, 2013	M
359	469	Jacqueline Kennedy Onassis sued him, Marlon Br...	PG-13	Documentary Musical and Performing Arts Specia...	Leon Gast	NaN	Jul 30, 2010	C
970	1260	With the recent assault on Berlusconi in Milan...	NR	Documentary Special Interest Television	Erik Gandini	Erik Gandini	Aug 28, 2009	

340 rows × 12 columns



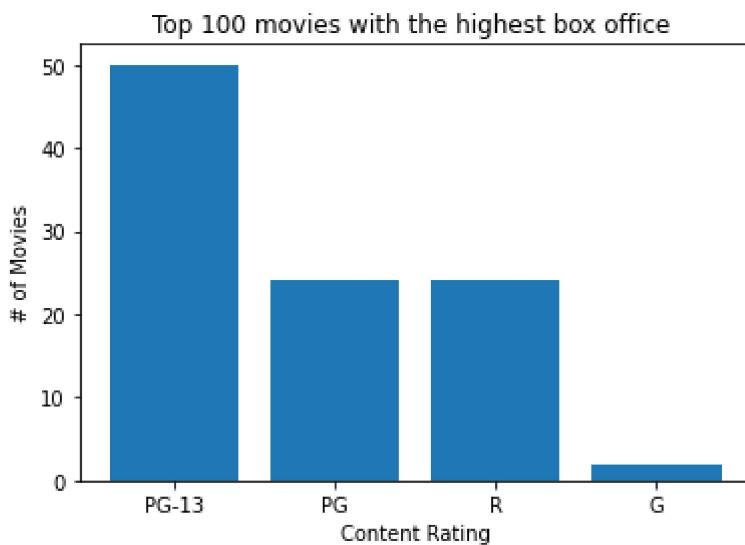
```
In [10]: bownv_top_100bo = bownv['rating'].head(100)
bownv_top_100bo.value_counts() # For the top 100 movies that have non-null box_office v
# Looks Like PG-13 appears the most.
```

```
Out[10]: PG-13    50  
PG        24  
R         24  
G         2  
Name: rating, dtype: int64
```

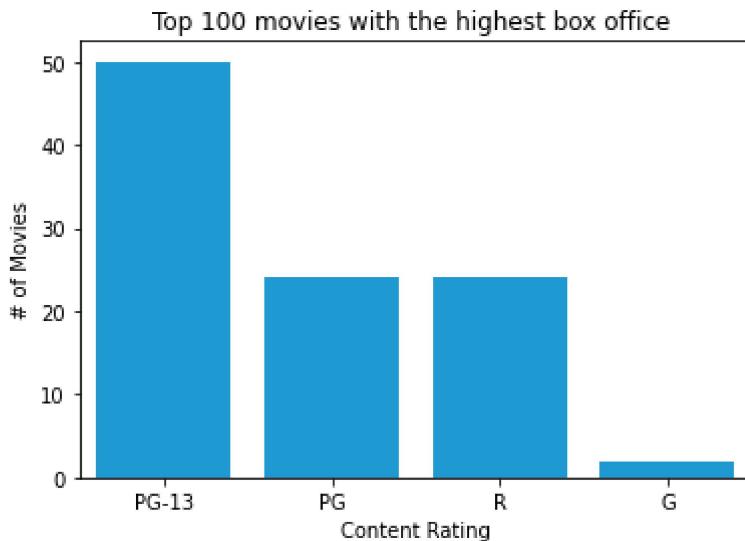
```
In [11]: bownv_top_100bo.value_counts().keys()
```

```
Out[11]: Index(['PG-13', 'PG', 'R', 'G'], dtype='object')
```

```
In [24]: fig,ax = plt.subplots()  
x = bownv_top_100bo.value_counts().keys()  
height = bownv_top_100bo.value_counts().values  
  
ax.bar( x = x , height = height);  
ax.set_xlabel('Content Rating')  
ax.set_ylabel('# of Movies')  
ax.set_title('Top 100 movies with the highest box office ');
```



```
In [85]: plot = sns.barplot(x = x , y = height, data = bownv_top_100bo , orient = 'v', color = 'red')  
plot.set_xlabel('Content Rating')  
plot.set_ylabel('# of Movies')  
plot.set_title('Top 100 movies with the highest box office ');
```

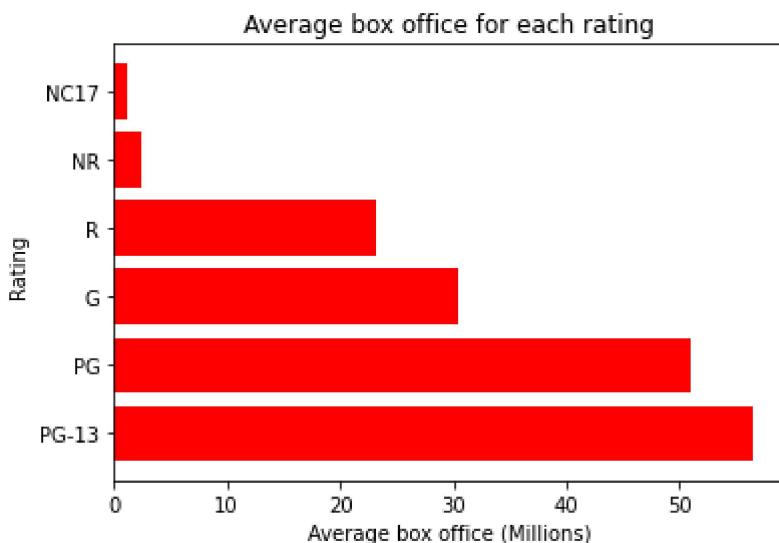


```
In [116... df_rating_avg_bf=bownv.groupby(by='rating').mean().reset_index()[['rating','box_office']]
df_rating_avg_bf.sort_values(by='box_office', ascending = False , inplace = True)

In [117... df_rating_avg_bf.box_office = df_rating_avg_bf.box_office / 1000000 # make it simple

In [118... fig, ax = plt.subplots()
ax.ticklabel_format(style='plain')
x = df_rating_avg_bf.rating
height = df_rating_avg_bf.box_office

ax.barh( y = x , width = height , color = 'red')
ax.set_xlabel('Average box office (Millions)')
ax.set_ylabel('Rating ')
ax.set_title('Average box office for each rating');
```



```
In [119... x = df_rating_avg_bf.rating
height = df_rating_avg_bf.box_office

In [122... df_rating_avg_bf.sort_values('box_office', ascending = True, inplace = True)

In [126... plot = sns.barplot(x = height , y = x, data = df_rating_avg_bf , orient = 'h', color =
plot.set_xlabel('Revenue (Millions)')
plot.set_ylabel('Content Rating')
plot.set_title('Average Box Office Revenue By Content Rating');
plt.savefig("../Images/Average Box Office Revenue By Content Rating.jpg", dpi=400, bbox
```

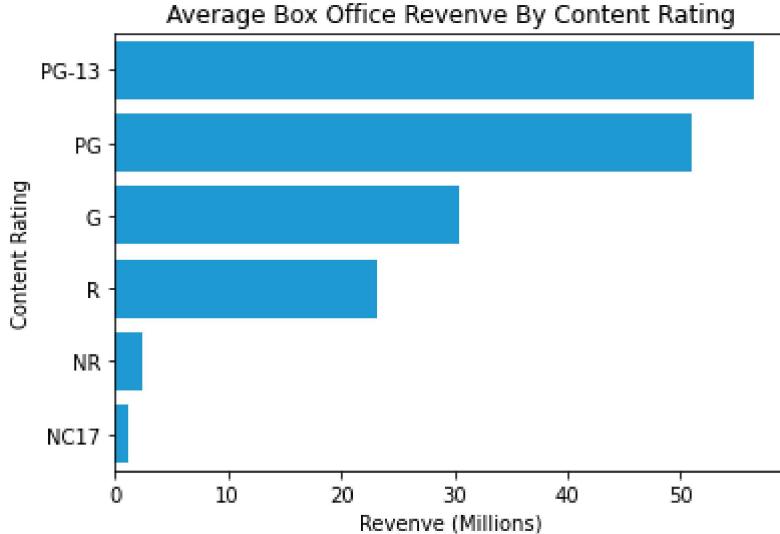
```
-----
FileNotFoundError Traceback (most recent call last)
<ipython-input-126-f2af256704d1> in <module>
      3 plot.set_ylabel('Content Rating')
      4 plot.set_title('Average Box Office Revenue By Content Rating');
----> 5 plt.savefig("../Images/Average Box Office Revenue By Content Rating.jpg", dpi=4
00, bbox_inches='tight')
```

```
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/pyplot.py in savefig(*args, **kwargs)
    841 def savefig(*args, **kwargs):
    842     fig = gcf()
--> 843     res = fig.savefig(*args, **kwargs)
    844     fig.canvas.draw_idle()  # need this if 'transparent=True' to reset colors
    845     return res
```

```
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/figure.py in savefig(self, fname, transparent, **kwargs)
    2309                 patch.set_edgecolor('none')
    2310
-> 2311         self.canvas.print_figure(fname, **kwargs)
    2312
    2313     if transparent:
...
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/backend_bases.py in print_figure(self, filename, dpi, facecolor, edgecolor, orientation, format, bbox_inches, pad_inches, bbox_extra_artists, backend, **kwargs)
    2208
    2209         try:
-> 2210             result = print_method(
    2211                 filename,
    2212                 dpi=dpi,
...
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/backend_bases.py in wrapper(*args, **kwargs)
    1637             kwargs.pop(arg)
    1638
-> 1639         return func(*args, **kwargs)
    1640
    1641     return wrapper
...
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/cbook/deprecation.py in wrapper(*inner_args, **inner_kwargs)
    409                     else deprecation_addendum,
    410                     **kwargs)
--> 411         return func(*inner_args, **inner_kwargs)
    412
    413     return wrapper
...
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/cbook/deprecation.py in wrapper(*inner_args, **inner_kwargs)
    409                     else deprecation_addendum,
    410                     **kwargs)
--> 411         return func(*inner_args, **inner_kwargs)
    412
    413     return wrapper
...
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/cbook/deprecation.py in wrapper(*inner_args, **inner_kwargs)
    409                     else deprecation_addendum,
    410                     **kwargs)
--> 411         return func(*inner_args, **inner_kwargs)
    412
    413     return wrapper
...
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/cbook/deprecation.py in wrapper(*inner_args, **inner_kwargs)
    409                     else deprecation_addendum,
    410                     **kwargs)
--> 411         return func(*inner_args, **inner_kwargs)
    412
    413     return wrapper
...
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/matplotlib/backends/backend_agg.py in print_jpg(self, filename_or_obj, dryrun, pil_kwds, *args, **kwargs)
    590         pil_kwds.setdefault("dpi", (self.figure.dpi, self.figure.dpi))
    591         # Drop alpha channel now.
--> 592         return (Image.fromarray(np.asarray(self.buffer_rgba())[..., :3])
    593                 .save(filename_or_obj, format='jpeg', **pil_kwds))
    594
```

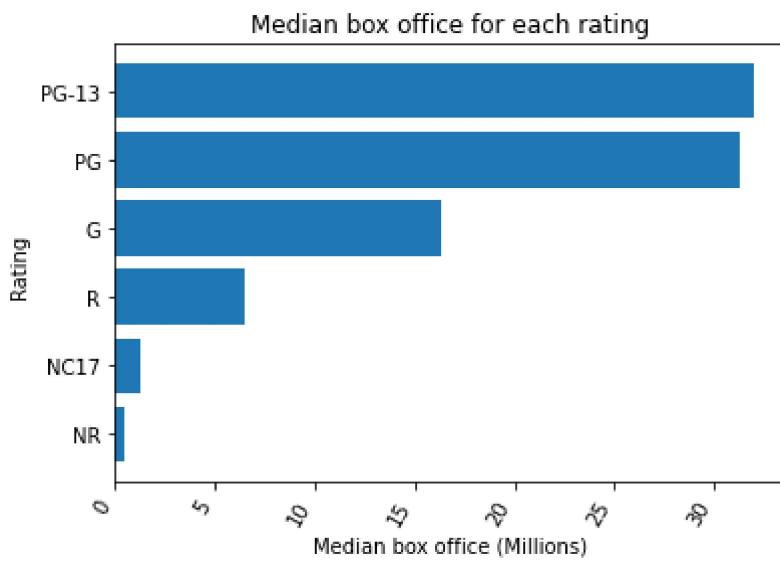
```
~/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/PIL/Image.py in save(self, f
p, format, **params)
    2153         fp = builtins.open(filename, "r+b")
    2154     else:
-> 2155         fp = builtins.open(filename, "w+b")
    2156
    2157     try:
```

`FileNotFoundException: [Errno 2] No such file or directory: '../Images/Average Box Office Revenue By Content Rating.jpg'`



```
In [103...]: df_rating_median_bf=bownv.groupby(by='rating').median().reset_index()[['rating','box_of
ice_office']]
In [52]: df_rating_median_bf.sort_values(by='box_office', ascending = True , inplace = True)
In [53]: df_rating_median_bf.box_office = df_rating_median_bf.box_office / 1000000
In [55]: fig, ax = plt.subplots()
ax.ticklabel_format(style='plain')
plt.xticks(rotation= 60, ha="right")
x = df_rating_median_bf.rating
height = df_rating_median_bf.box_office

ax.barh( y = x , width = height);
ax.set_xlabel('Median box office (Millions)')
ax.set_ylabel('Rating ')
ax.set_title('Median box office for each rating');
```



```
In [19]: # From the information given above, movies that rated PG-13 and PG have higher box offi
```

```
In [ ]:
```