

```
In [67]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

```
In [68]: movie_dataset = pd.read_csv(r'C:\Users\bmittipa\Documents\Vodafone\PrakashSenapati\
```

```
In [69]: movie_dataset.head()
```

```
Out[69]:
```

	Film	Genre	Rotten Tomatoes Ratings %	Audience Ratings %	Budget (million \$)	Year of release
0	(500) Days of Summer	Comedy	87	81	8	2009
1	10,000 B.C.	Adventure	9	44	105	2008
2	12 Rounds	Action	30	52	20	2009
3	127 Hours	Adventure	93	84	18	2010
4	17 Again	Comedy	55	70	20	2009

```
In [70]: type(movie_dataset)
```

```
Out[70]: pandas.core.frame.DataFrame
```

```
In [71]: len(movie_dataset)
```

```
Out[71]: 559
```

```
In [72]: movie_dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 559 entries, 0 to 558
Data columns (total 6 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Film                                  559 non-null    object
1   Genre                                559 non-null    object
2   Rotten Tomatoes Ratings %            559 non-null    int64
3   Audience Ratings %                   559 non-null    int64
4   Budget (million $)                   559 non-null    int64
5   Year of release                       559 non-null    int64
dtypes: int64(4), object(2)
memory usage: 26.3+ KB
```

```
In [73]: print(np.__version__)
print(pd.__version__)
```

1.26.4
2.2.2

```
In [74]: movie_dataset.columns
```

```
Out[74]: Index(['Film', 'Genre', 'Rotten Tomatoes Ratings %', 'Audience Ratings %',  
              'Budget (million $)', 'Year of release'],  
              dtype='object')
```

```
In [75]: movie_dataset.shape
```

```
Out[75]: (559, 6)
```

```
In [76]: movie_dataset.head()
```

Out[76]:

	Film	Genre	Rotten Tomatoes Ratings %	Audience Ratings %	Budget (million \$)	Year of release
0	(500) Days of Summer	Comedy	87	81	8	2009
1	10,000 B.C.	Adventure	9	44	105	2008
2	12 Rounds	Action	30	52	20	2009
3	127 Hours	Adventure	93	84	18	2010
4	17 Again	Comedy	55	70	20	2009

```
In [77]: movie_dataset.columns
```

```
Out[77]: Index(['Film', 'Genre', 'Rotten Tomatoes Ratings %', 'Audience Ratings %',  
              'Budget (million $)', 'Year of release'],  
              dtype='object')
```

```
In [78]: movie_dataset.columns = ['Film','Genre','CriticRating','AudienceRating','BudgetMill
```

```
In [79]: movie_dataset.head()
```

Out[79]:

	Film	Genre	CriticRating	AudienceRating	BudgetMillions	Year
0	(500) Days of Summer	Comedy	87	81	8	2009
1	10,000 B.C.	Adventure	9	44	105	2008
2	12 Rounds	Action	30	52	20	2009
3	127 Hours	Adventure	93	84	18	2010
4	17 Again	Comedy	55	70	20	2009

```
In [80]: movie_dataset.describe()
```

Out[80]:

	CriticRating	AudienceRating	BudgetMillions	Year
count	559.000000	559.000000	559.000000	559.000000
mean	47.309481	58.744186	50.236136	2009.152057
std	26.413091	16.826887	48.731817	1.362632
min	0.000000	0.000000	0.000000	2007.000000
25%	25.000000	47.000000	20.000000	2008.000000
50%	46.000000	58.000000	35.000000	2009.000000
75%	70.000000	72.000000	65.000000	2010.000000
max	97.000000	96.000000	300.000000	2011.000000

In [81]: `movie_dataset.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 559 entries, 0 to 558
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Film            559 non-null   object
1   Genre           559 non-null   object
2   CriticRating    559 non-null   int64
3   AudienceRating  559 non-null   int64
4   BudgetMillions  559 non-null   int64
5   Year            559 non-null   int64
dtypes: int64(4), object(2)
memory usage: 26.3+ KB
```

In []:

```
In [82]: movie_dataset.Film = movie_dataset.Film.astype('category')
movie_dataset.Genre = movie_dataset.Genre.astype('category')
movie_dataset.Year = movie_dataset.Year.astype('category')
```

In [83]: `movie_dataset.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 559 entries, 0 to 558
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Film            559 non-null   category
1   Genre           559 non-null   category
2   CriticRating    559 non-null   int64
3   AudienceRating  559 non-null   int64
4   BudgetMillions  559 non-null   int64
5   Year            559 non-null   category
dtypes: category(3), int64(3)
memory usage: 36.5 KB
```

In [84]: `movie_dataset.Genre`

Out[84]:

```

0      Comedy
1    Adventure
2      Action
3    Adventure
4      Comedy
...
554    Comedy
555    Comedy
556  Thriller
557    Action
558    Comedy
Name: Genre, Length: 559, dtype: category
Categories (7, object): ['Action', 'Adventure', 'Comedy', 'Drama', 'Horror', 'Romance', 'Thriller']

```

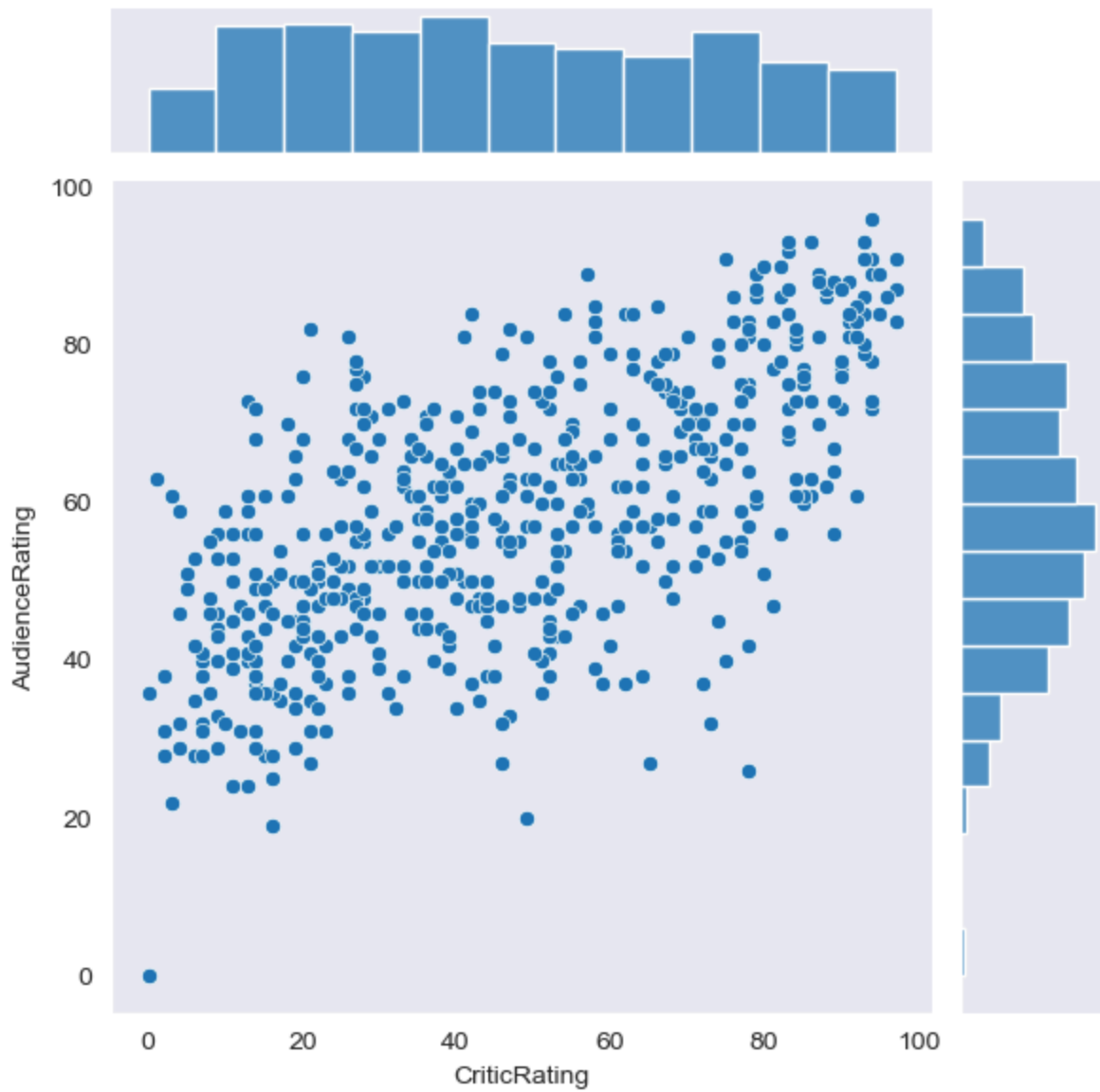
In [85]: `movie_dataset.head()`

Out[85]:

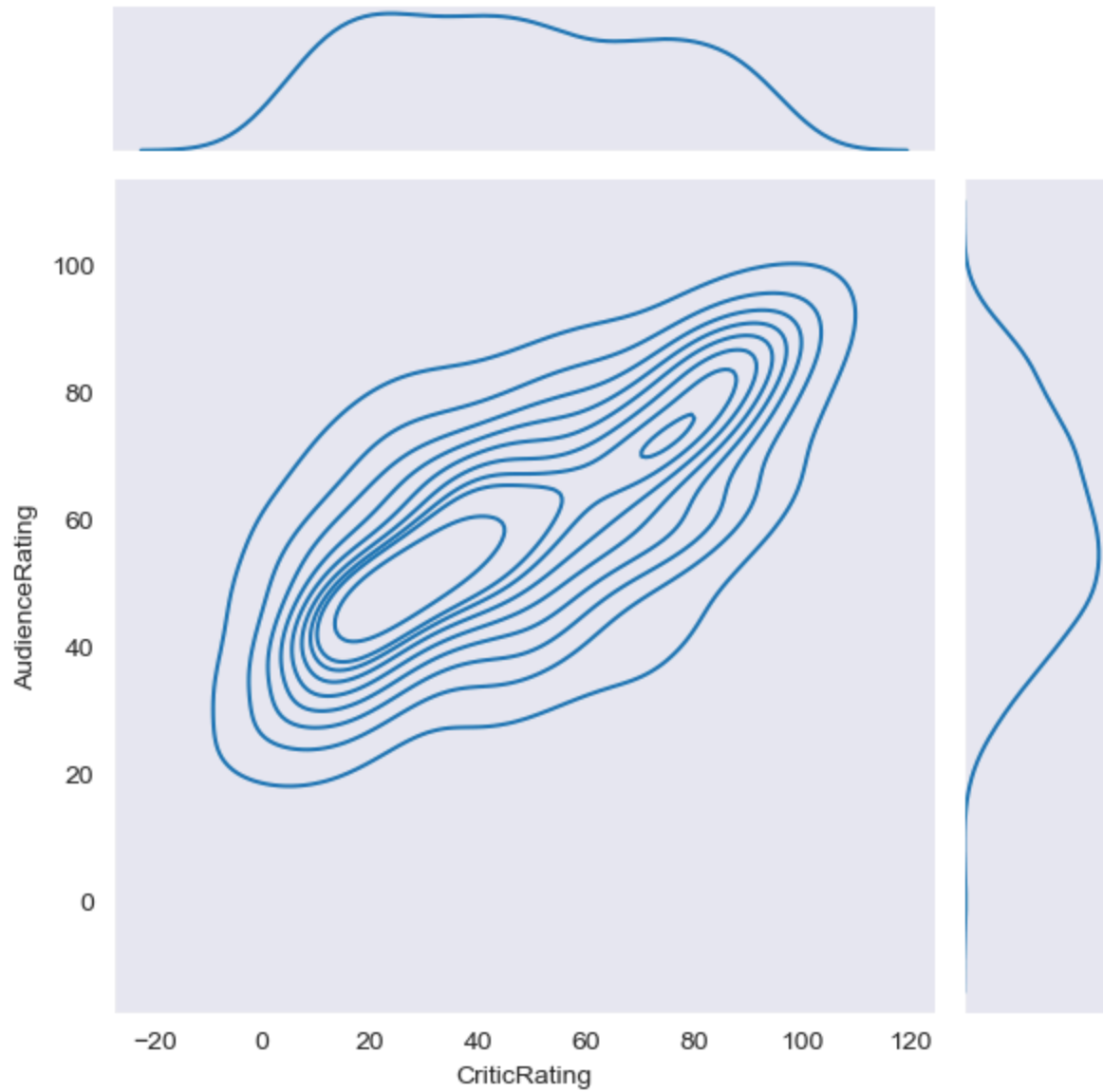
	Film	Genre	CriticRating	AudienceRating	BudgetMillions	Year
0	(500) Days of Summer	Comedy	87	81	8	2009
1	10,000 B.C.	Adventure	9	44	105	2008
2	12 Rounds	Action	30	52	20	2009
3	127 Hours	Adventure	93	84	18	2010
4	17 Again	Comedy	55	70	20	2009

In [86]: `sns.jointplot(data=movie_dataset, x='CriticRating', y='AudienceRating')`
`plt.show()`

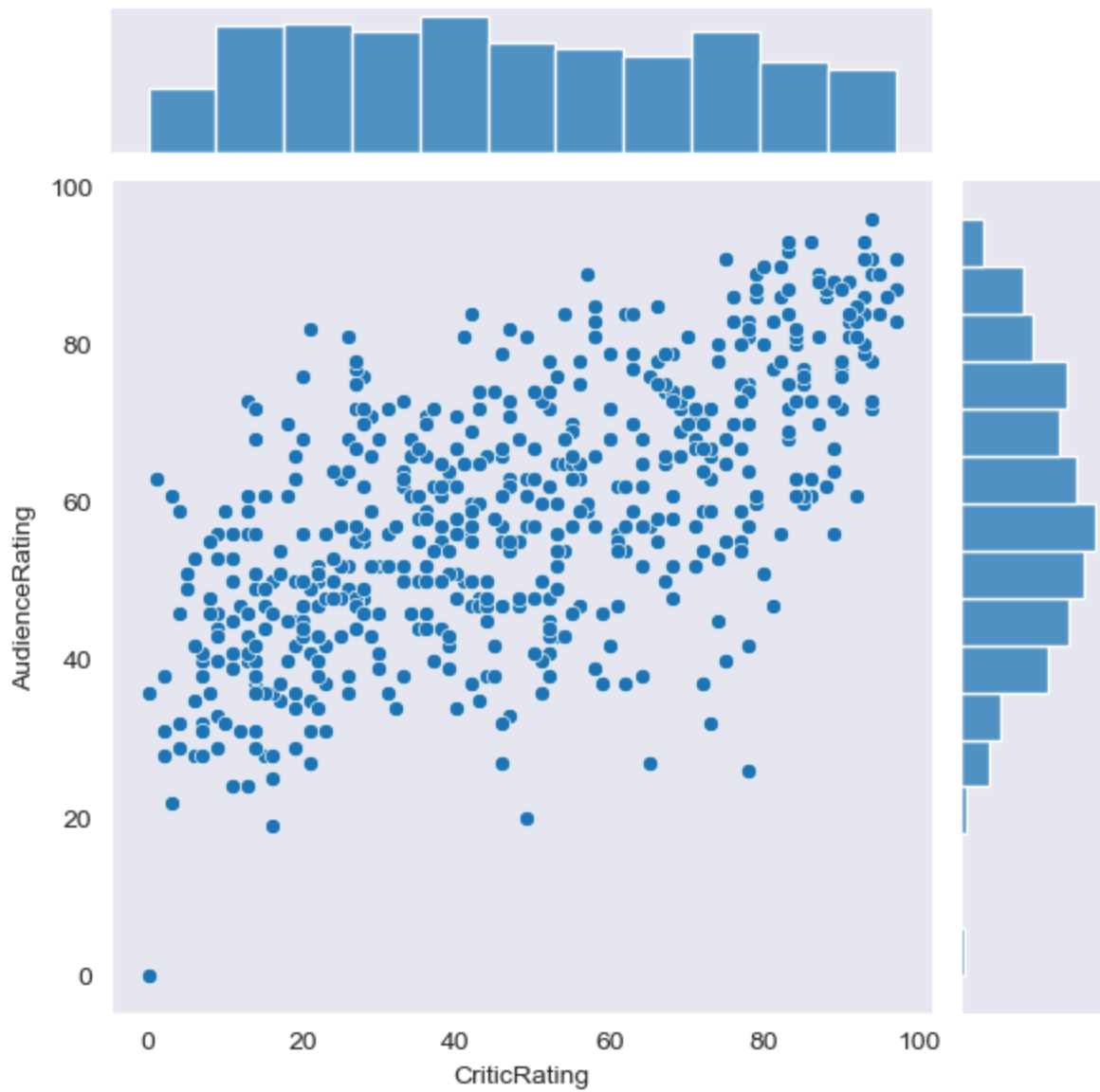




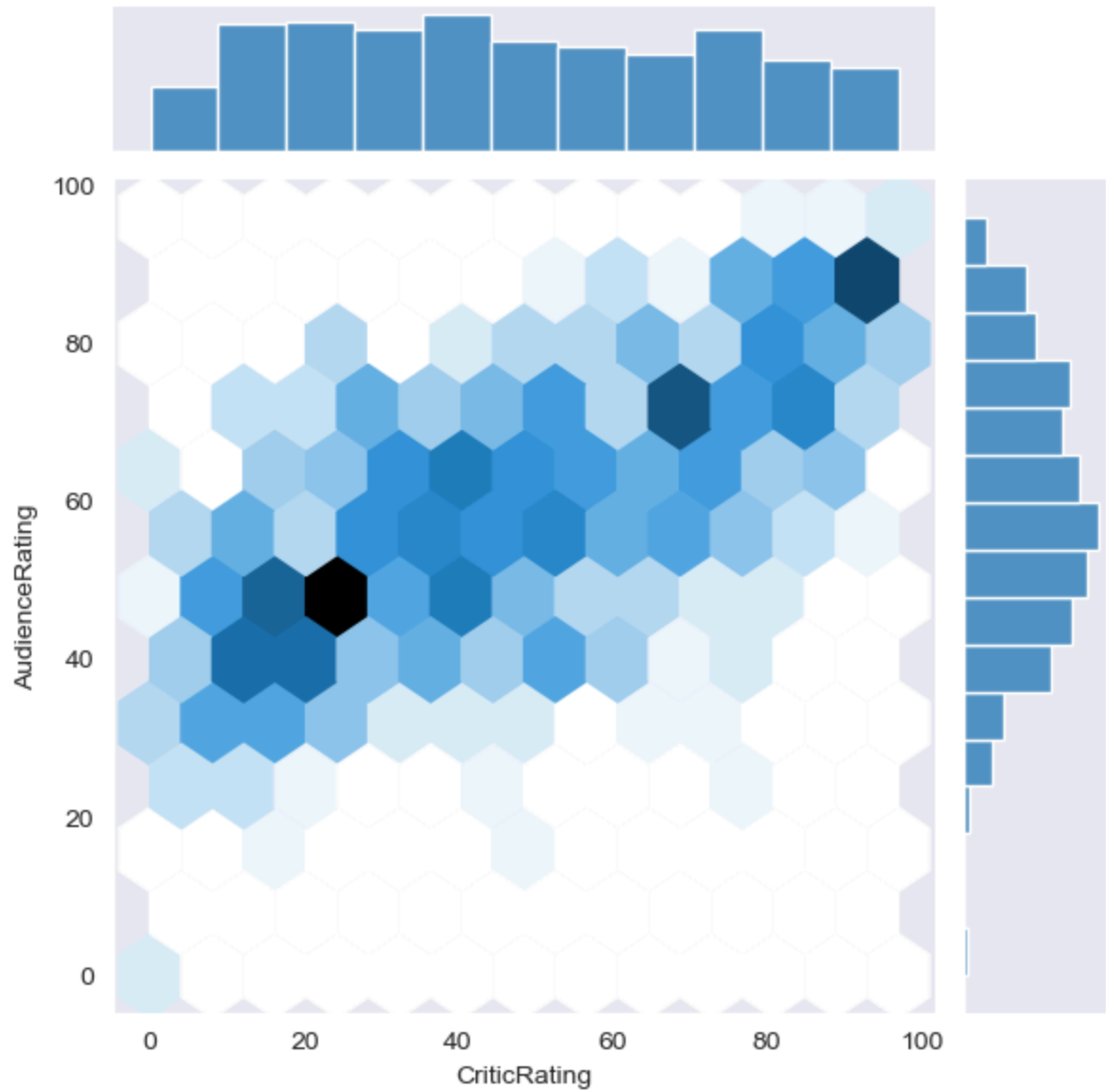
```
In [87]: sns.jointplot(data=movie_dataset, x='CriticRating',y='AudienceRating',kind='kde')  
plt.show()
```



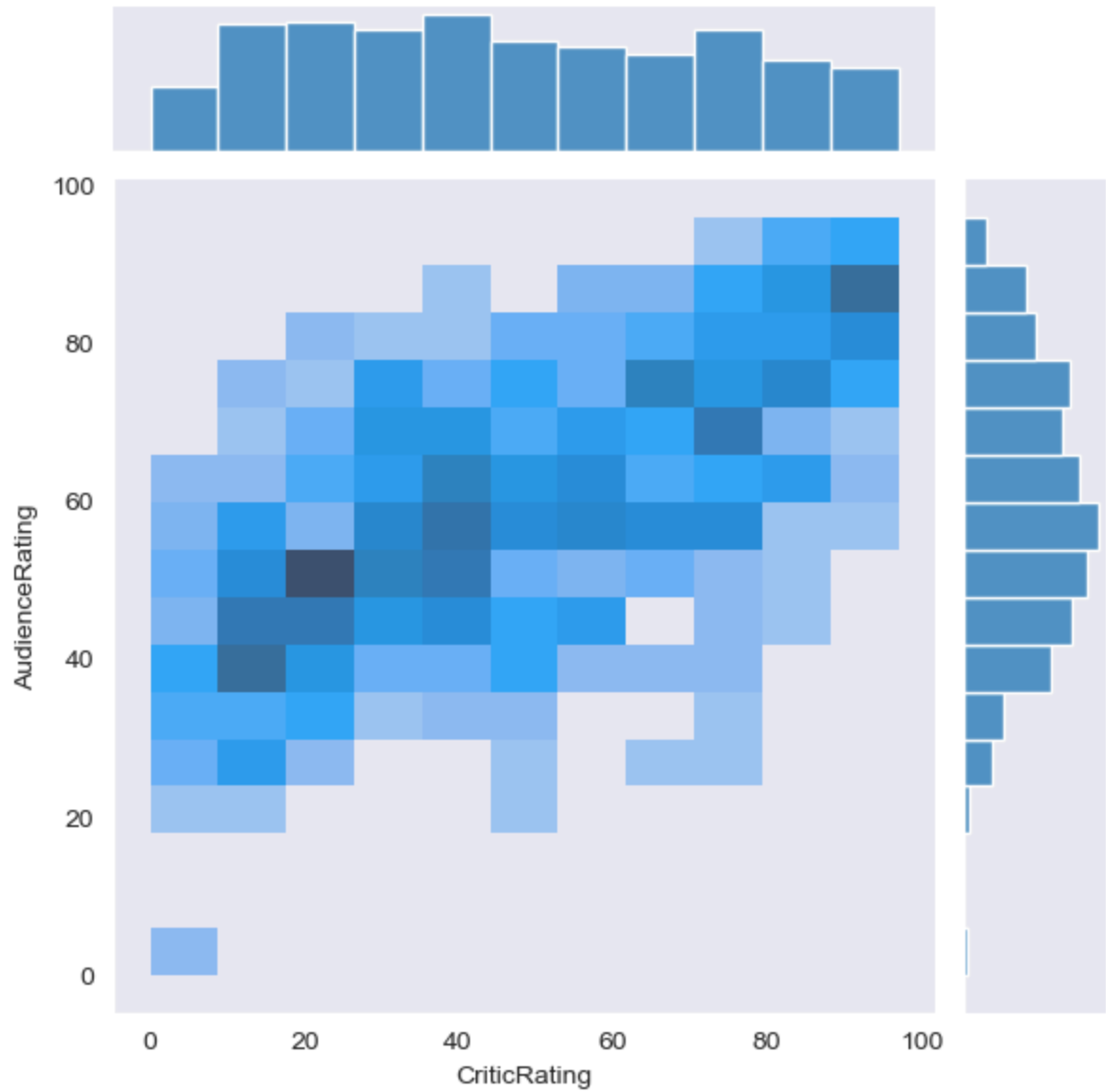
```
In [88]: sns.jointplot(data=movie_dataset, x='CriticRating',y='AudienceRating',kind='scatter',
plt.show())
```



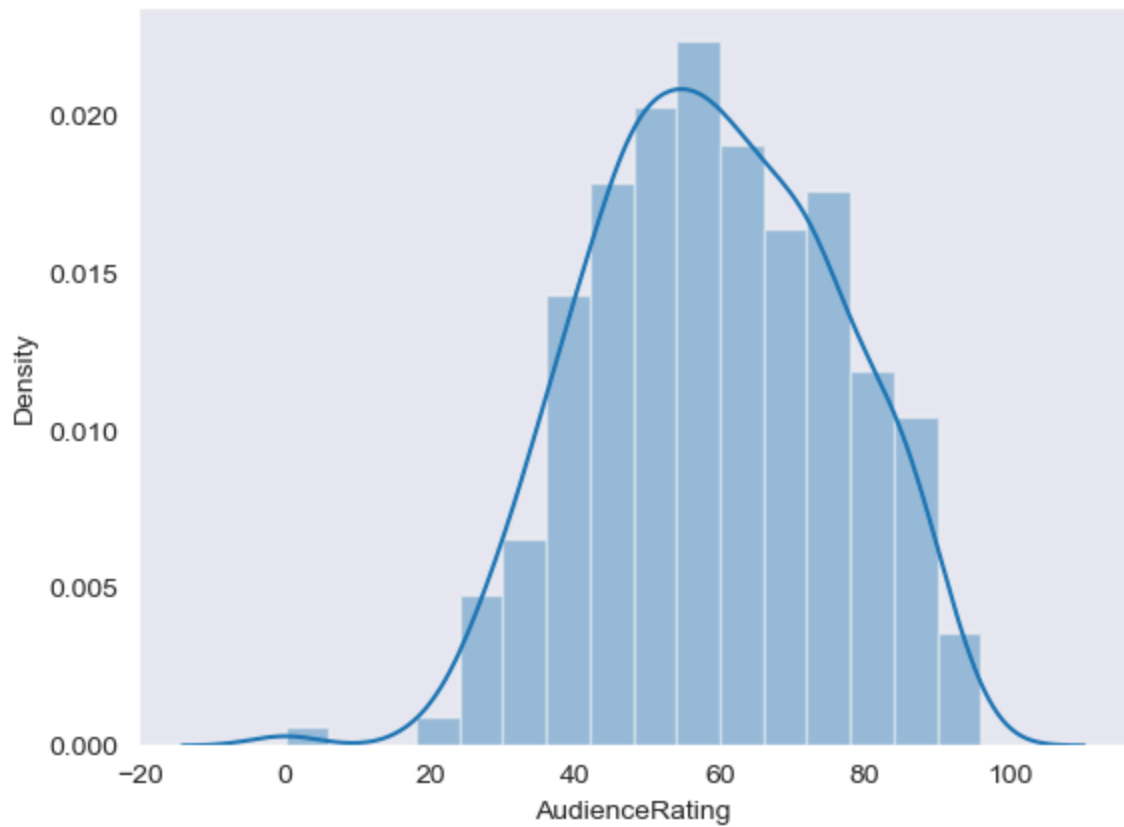
```
In [89]: sns.jointplot(data=movie_dataset, x='CriticRating',y='AudienceRating',kind='hex')  
plt.show()
```

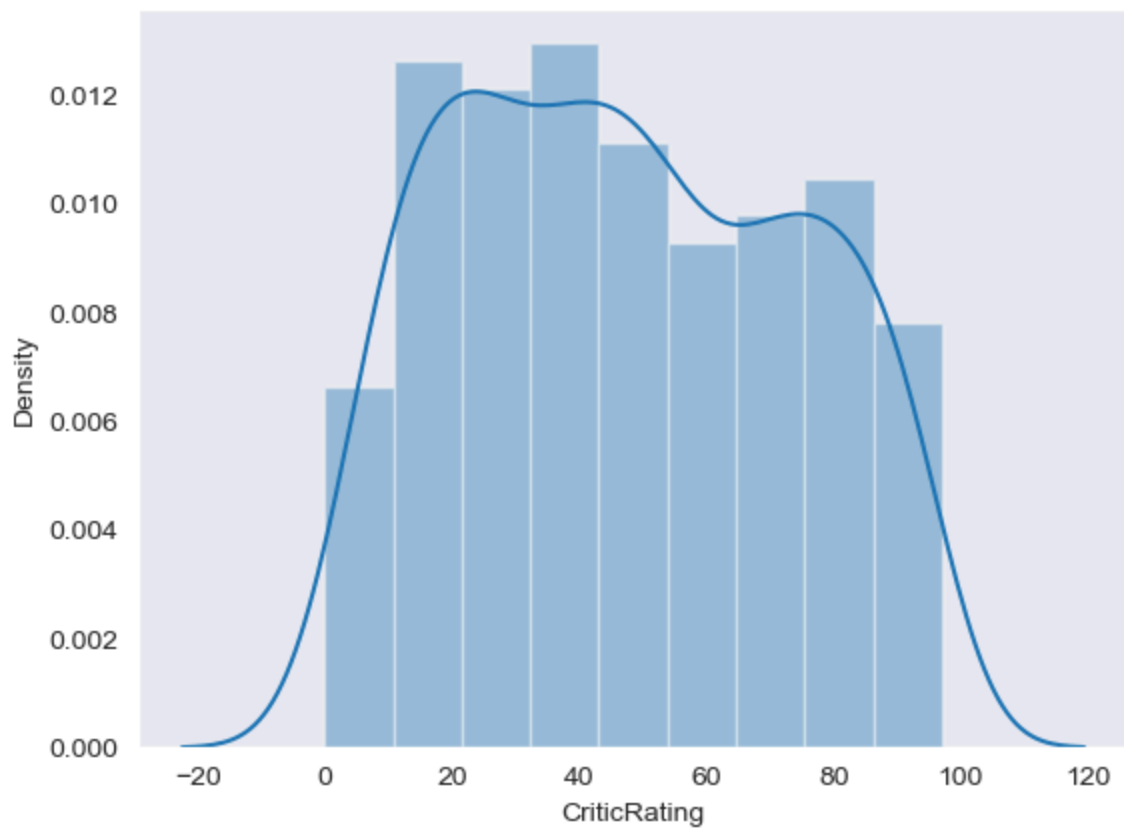
```
In [90]: sns.jointplot(data=movie_dataset, x='CriticRating',y='AudienceRating',kind='hist')
plt.show()
```



```
In [91]: sns.distplot(movie_dataset.AudienceRating)
plt.show()
```

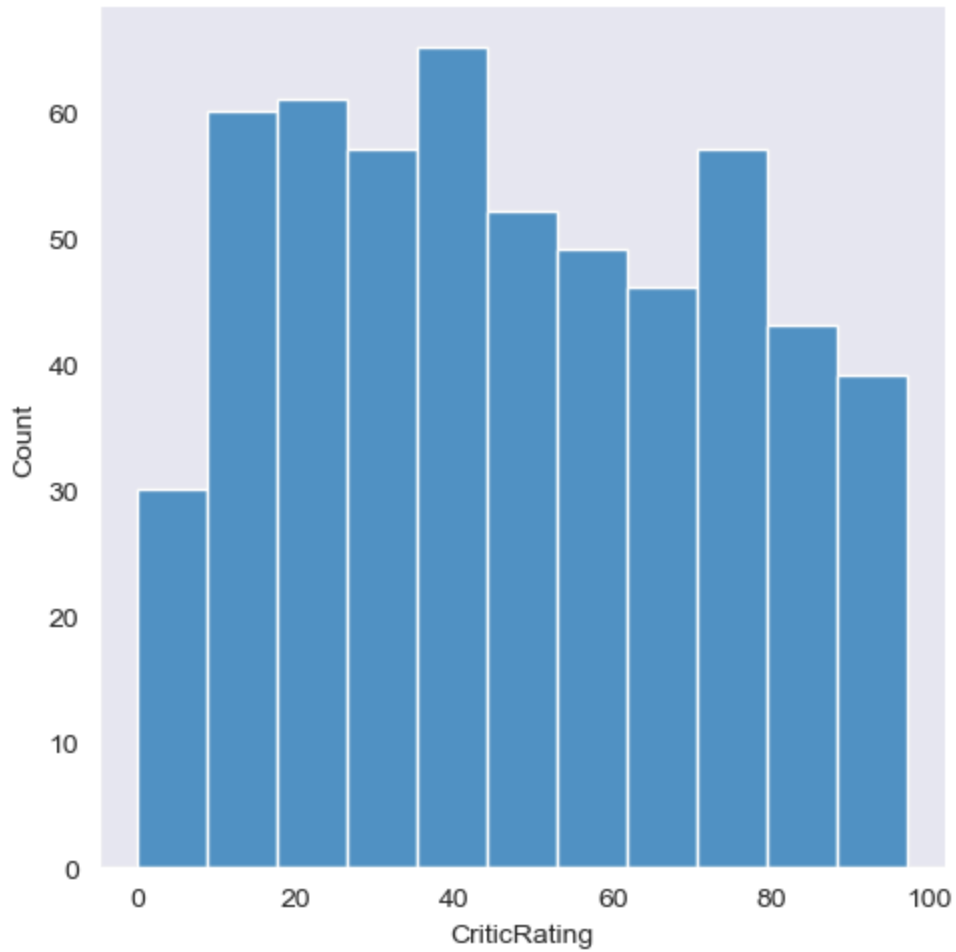


```
In [92]: sns.distplot(movie_dataset.CriticRating)
plt.show()
```

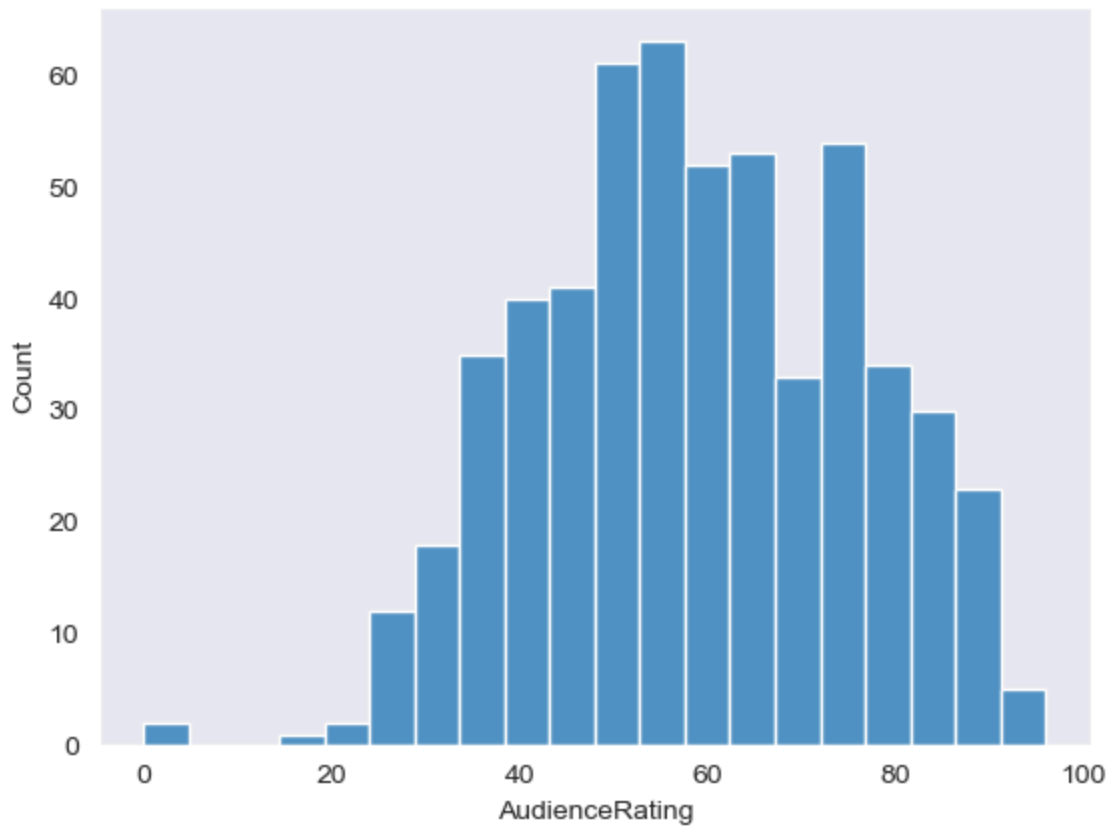


```
In [93]: sns.set_style('dark')
```

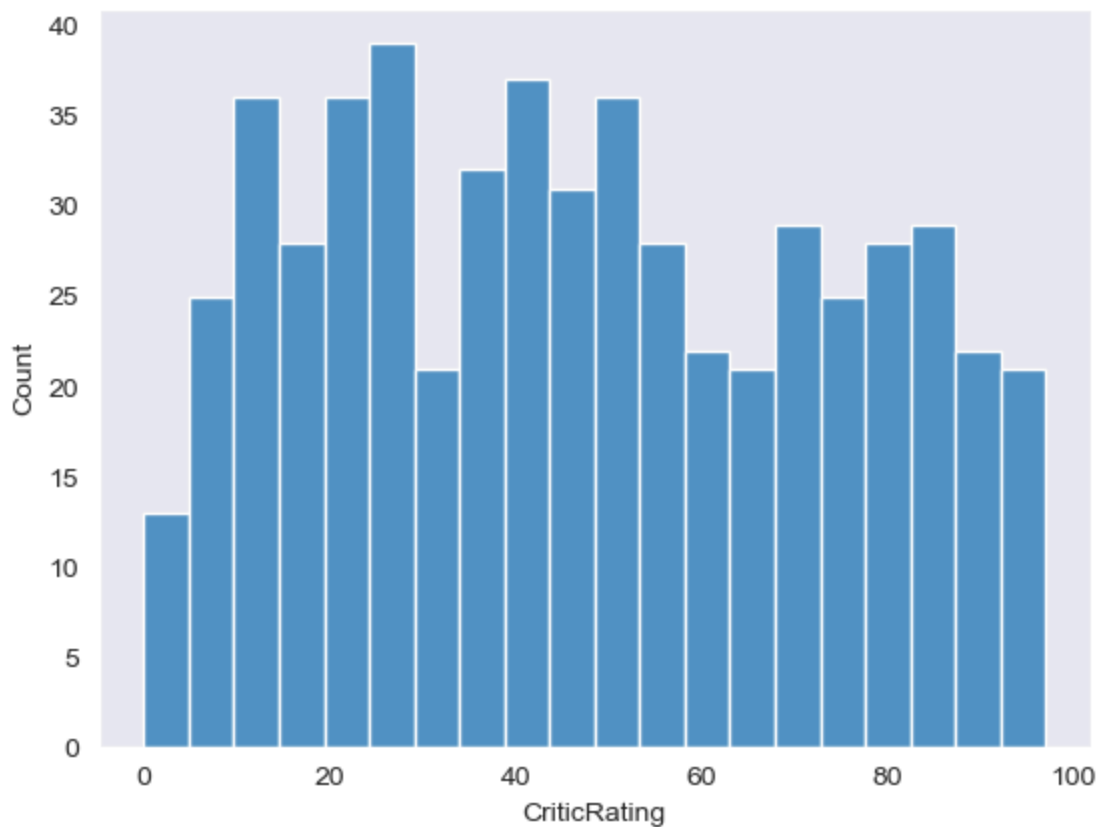
```
In [94]: sns.displot(movie_dataset.CriticRating)
plt.show()
```



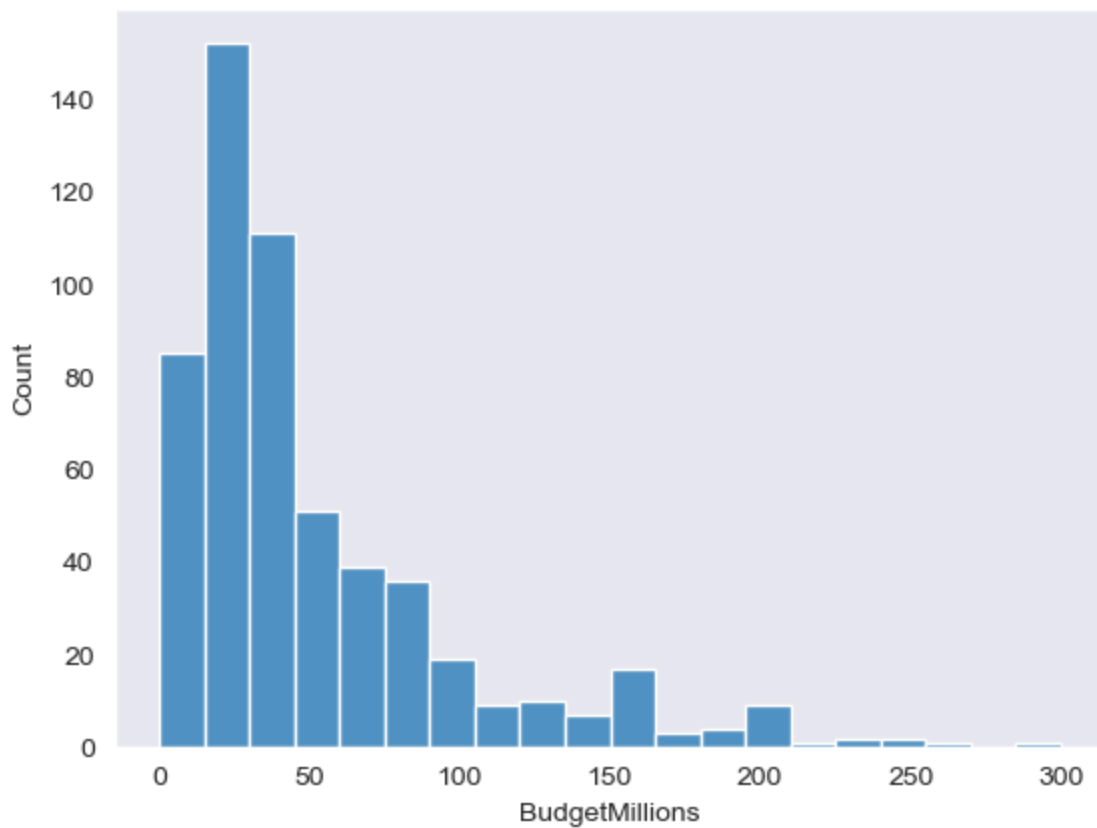
```
In [95]: sns.histplot(movie_dataset.AudienceRating, bins =20)
plt.show()
```



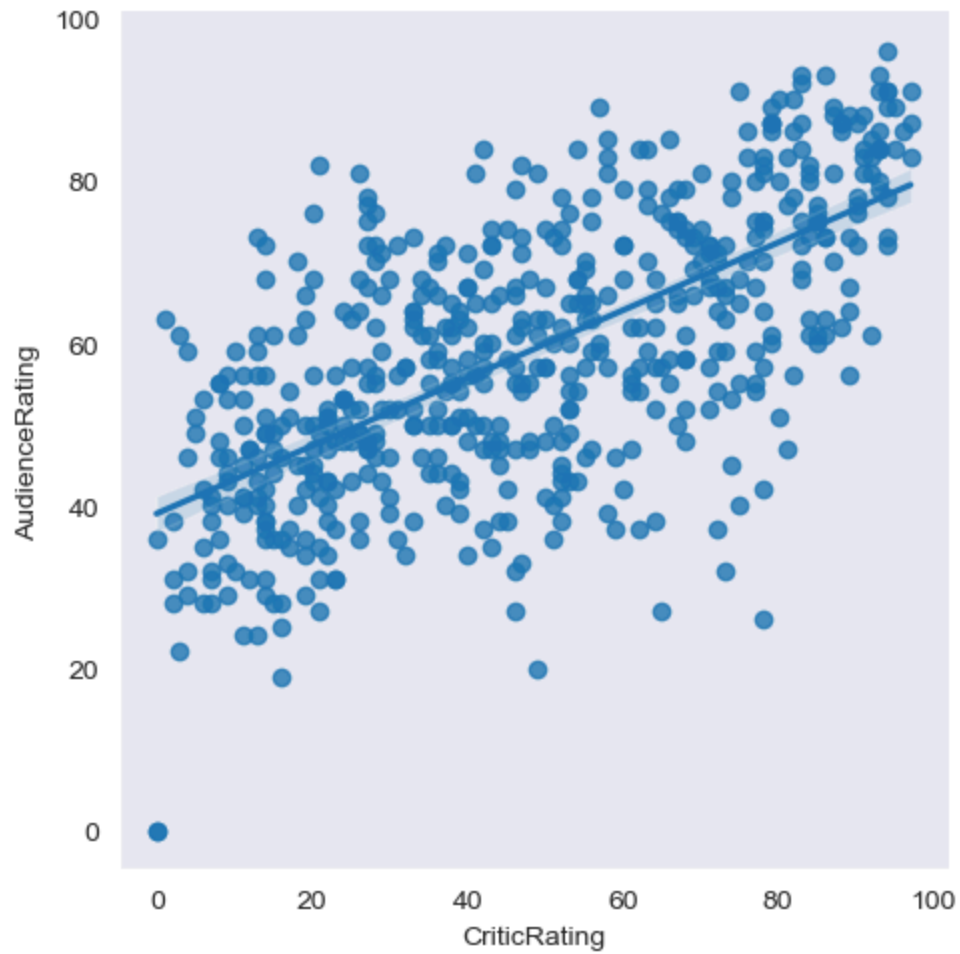
```
In [96]: sns.histplot(movie_dataset.CriticRating, bins =20)  
plt.show()
```



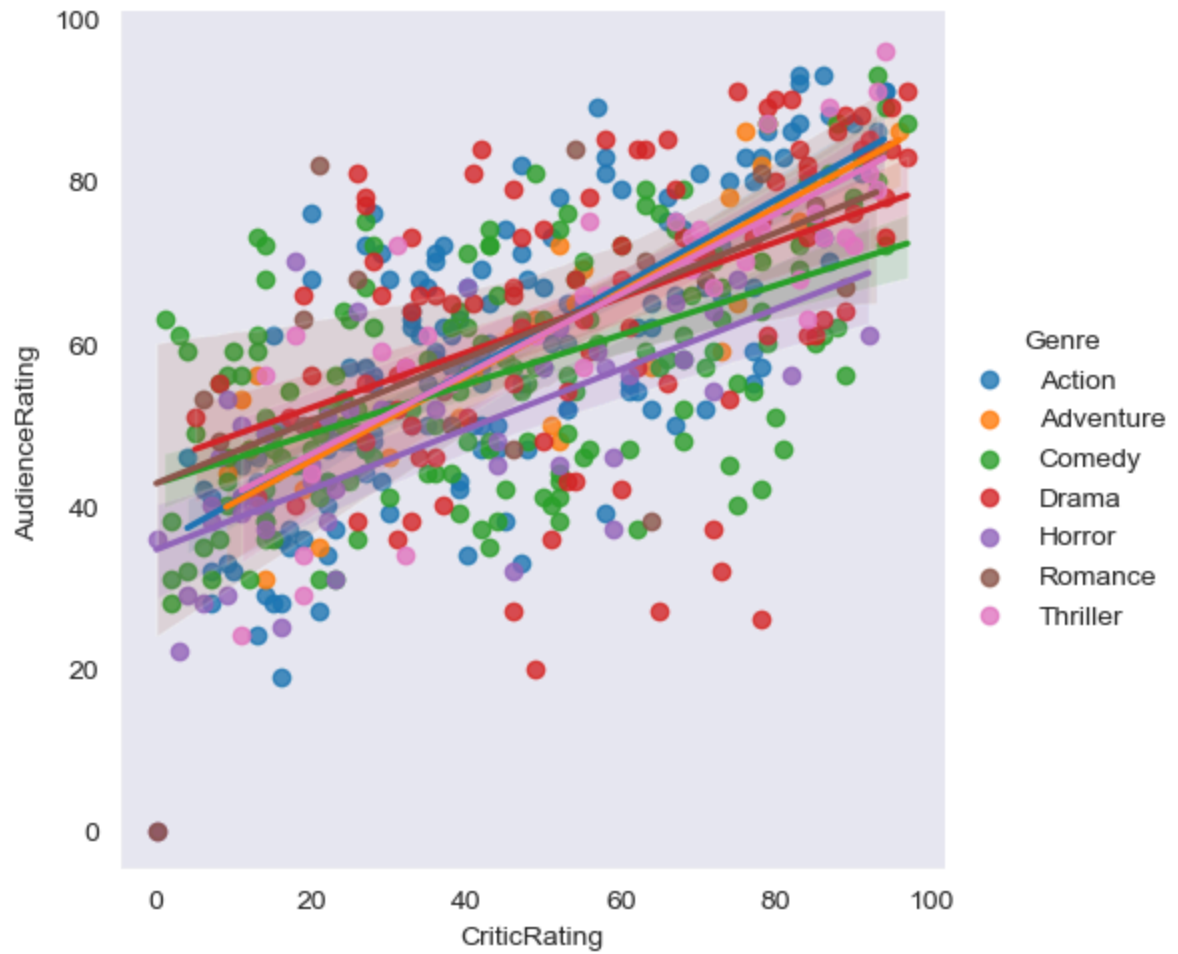
```
In [97]: sns.histplot(movie_dataset.BudgetMillions, bins =20)  
plt.show()
```



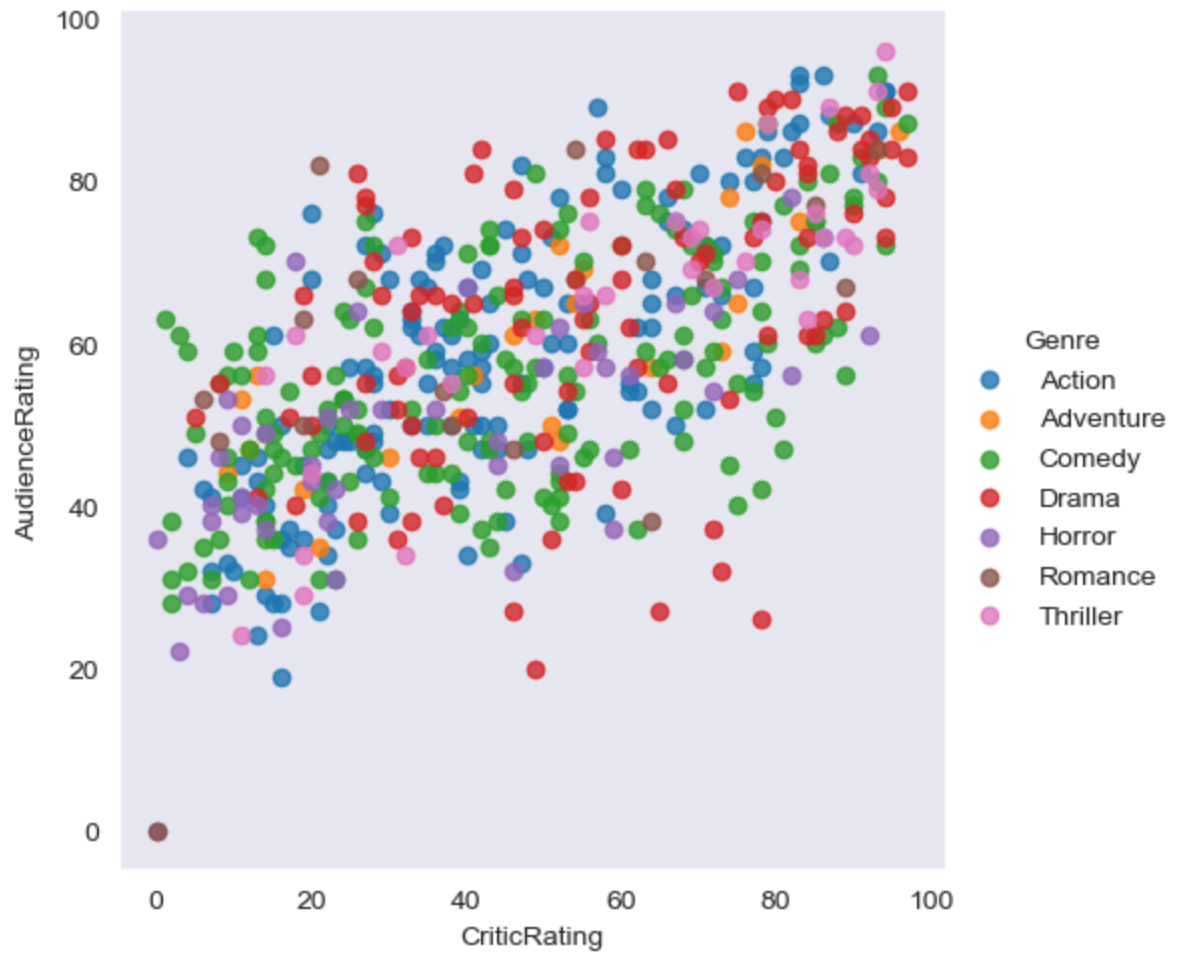
```
In [98]: sns.lmplot(data=movie_dataset, x='CriticRating',y='AudienceRating',fit_reg=True)  
plt.show()
```



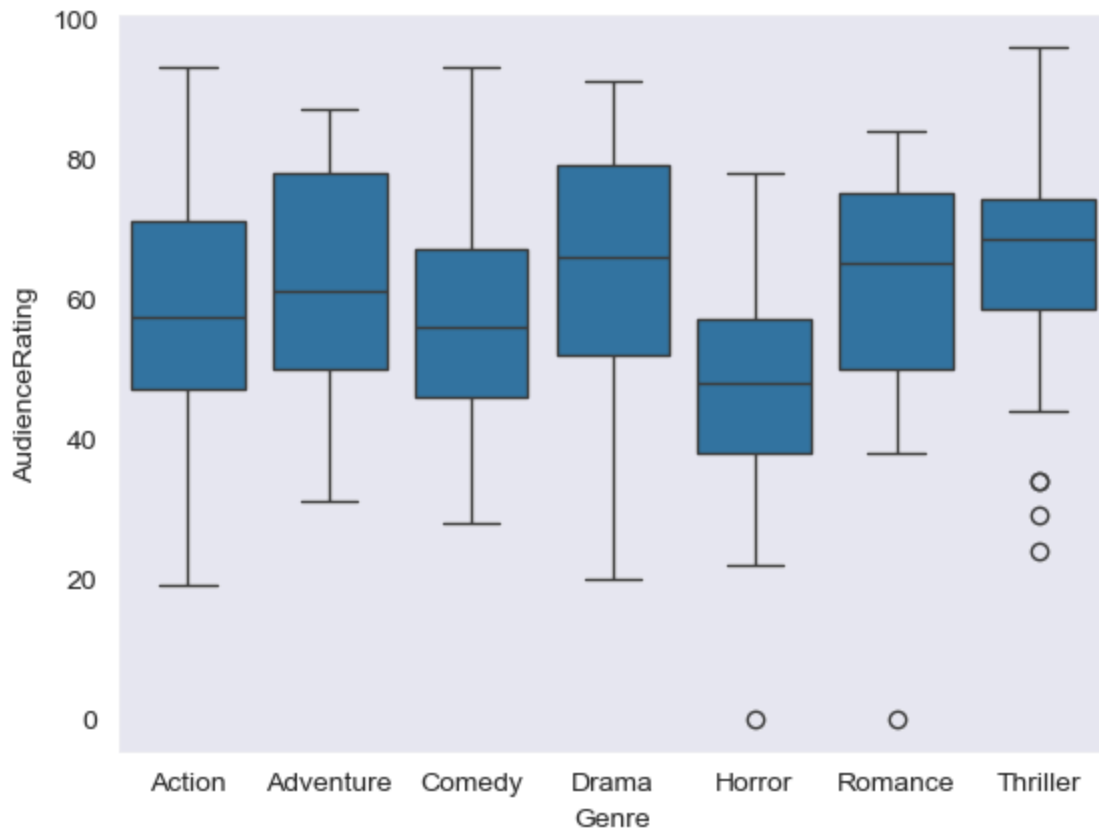
```
In [99]: sns.lmplot(data=movie_dataset, x='CriticRating', y='AudienceRating', fit_reg=True, hue=
plt.show()
```



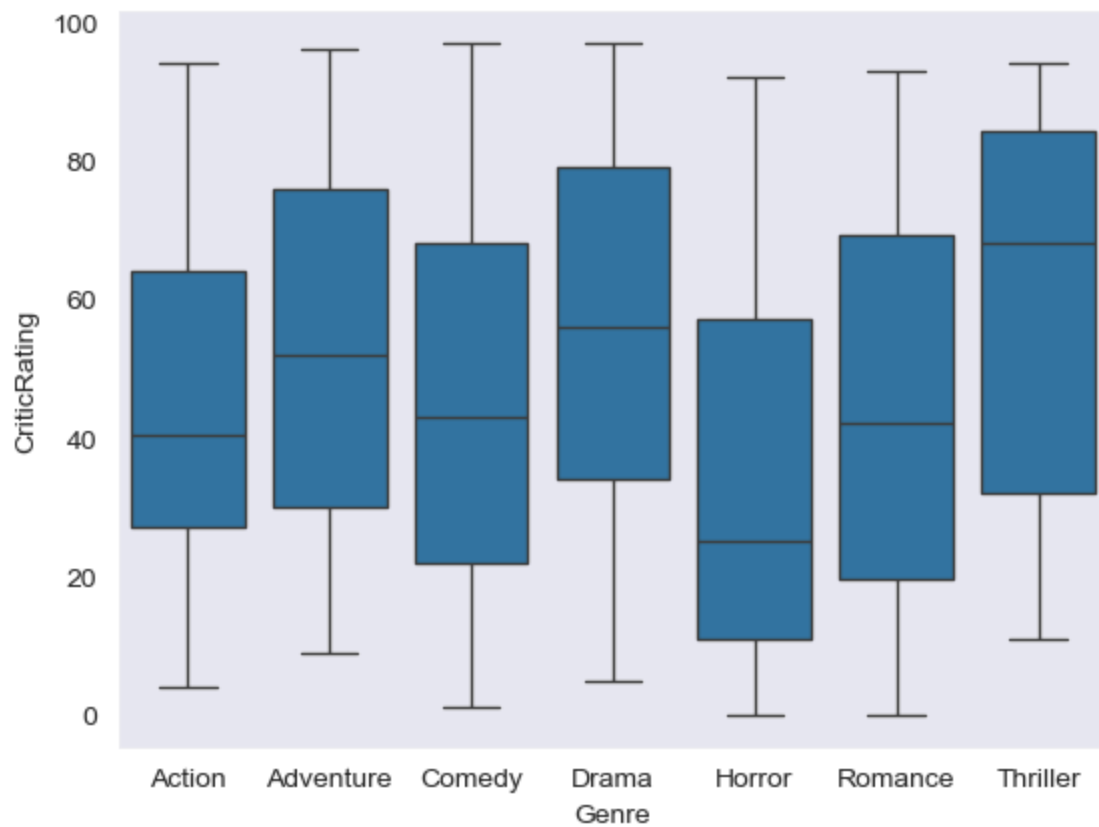
```
In [100... sns.lmplot(data=movie_dataset, x='CriticRating',y='AudienceRating',fit_reg=False,hu  
plt.show())
```

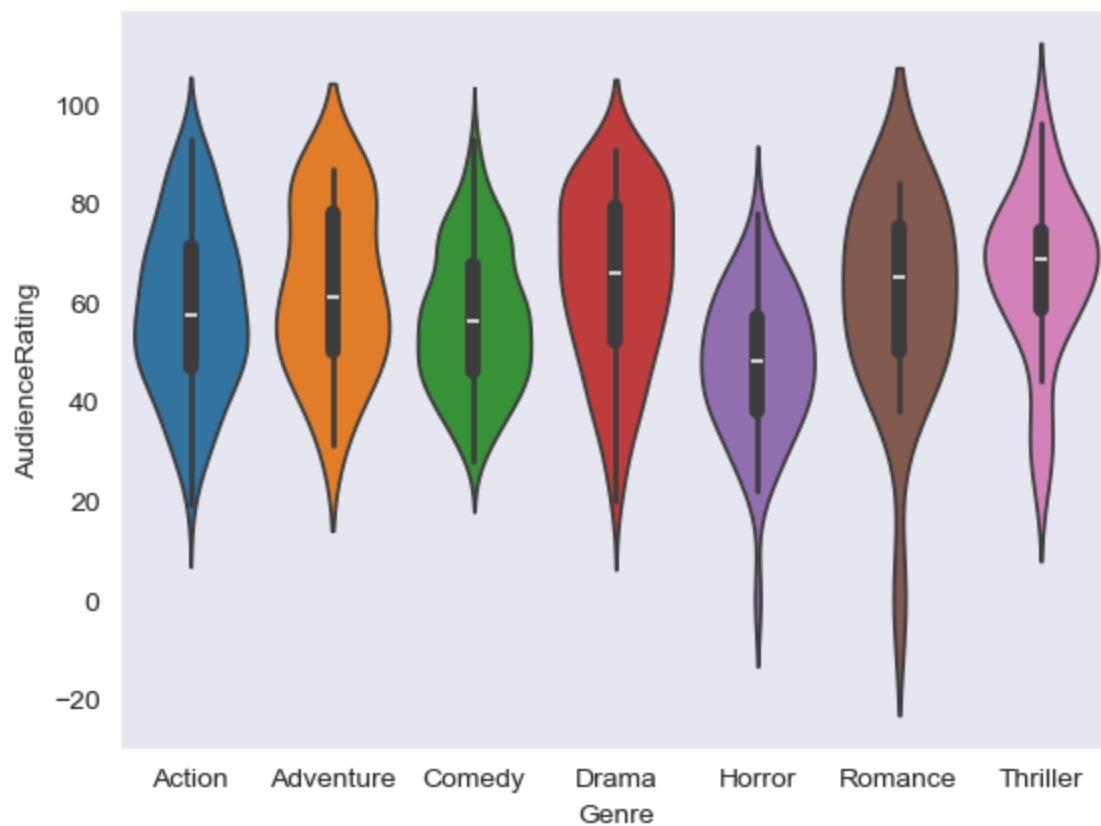
```
In [105... sns.boxplot(data=movie_dataset, x='Genre',y='AudienceRating')  
plt.show()
```



```
In [106... sns.boxplot(data=movie_dataset, x='Genre',y='CriticRating')  
plt.show()
```



```
In [108... sns.violinplot(data=movie_dataset, x='Genre',y='AudienceRating',hue='Genre')  
plt.show()
```



```
In [ ]:
```