In [1]:
```
import matplotlib.pyplot as plt
import pandas as pd
```

In [2]:
```
df = pd.read_csv('D:/data/movies.csv')
print(df.head())
```

```
   movie_id                   title     genre  release_year  imdb_rating  \
0         1               Inception    Sci-Fi          2010          8.8
1         2           KGF Chapter 2    Action          2022          8.4
2         3  Money Heist: The Movie  Thriller          2021          7.9
3         4                 Pathaan    Action          2023          6.9
4         5            Interstellar    Sci-Fi          2014          8.6

   duration_min      platform country  budget_million  box_office_million
0           148       Netflix     USA             160                 829
1           168  Amazon Prime   India              14                1230
2           131       Netflix   Spain              25                 210
3           146  Amazon Prime   India              30                1050
4           169       Netflix     USA             165                 677
```

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   movie_id            20 non-null     int64
 1   title               20 non-null     object
 2   genre               20 non-null     object
 3   release_year        20 non-null     int64
 4   imdb_rating         20 non-null     float64
 5   duration_min        20 non-null     int64
 6   platform            20 non-null     object
 7   country             20 non-null     object
 8   budget_million      20 non-null     int64
 9   box_office_million  20 non-null     int64
dtypes: float64(1), int64(5), object(4)
memory usage: 1.7+ KB
```

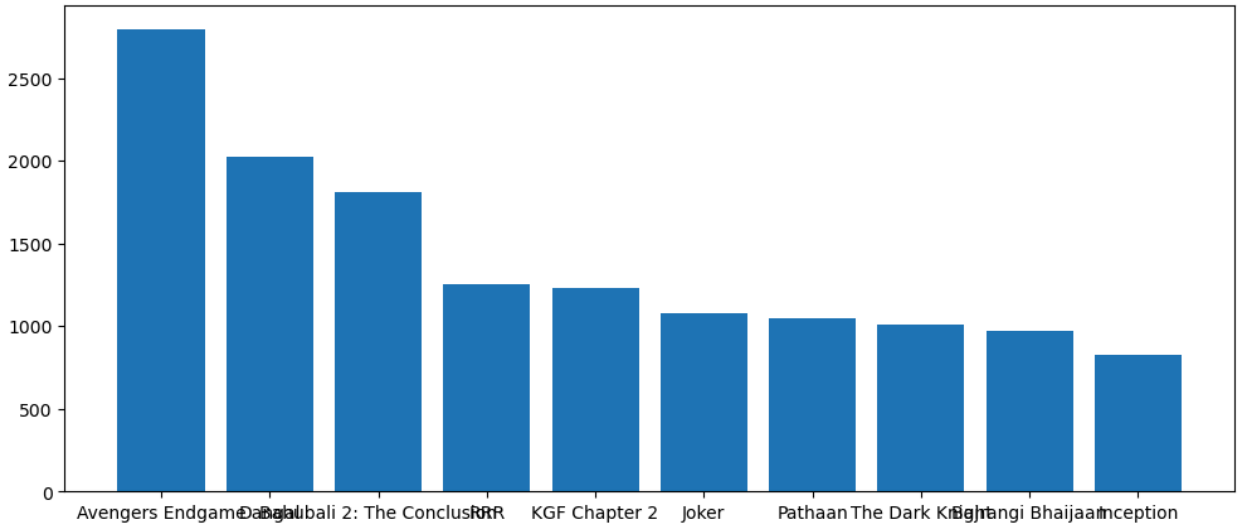
In [4]:
```
df.head()
```

Out[4]:

	movie_id	title	genre	release_year	imdb_rating	duration_min	platfor
0	1	Inception	Sci-Fi	2010	8.8	148	Netf
1	2	KGF Chapter 2	Action	2022	8.4	168	Amaz Prin
2	3	Money Heist: The Movie	Thriller	2021	7.9	131	Netf
3	4	Pathaan	Action	2023	6.9	146	Amaz Prin
4	5	Interstellar	Sci-Fi	2014	8.6	169	Netf

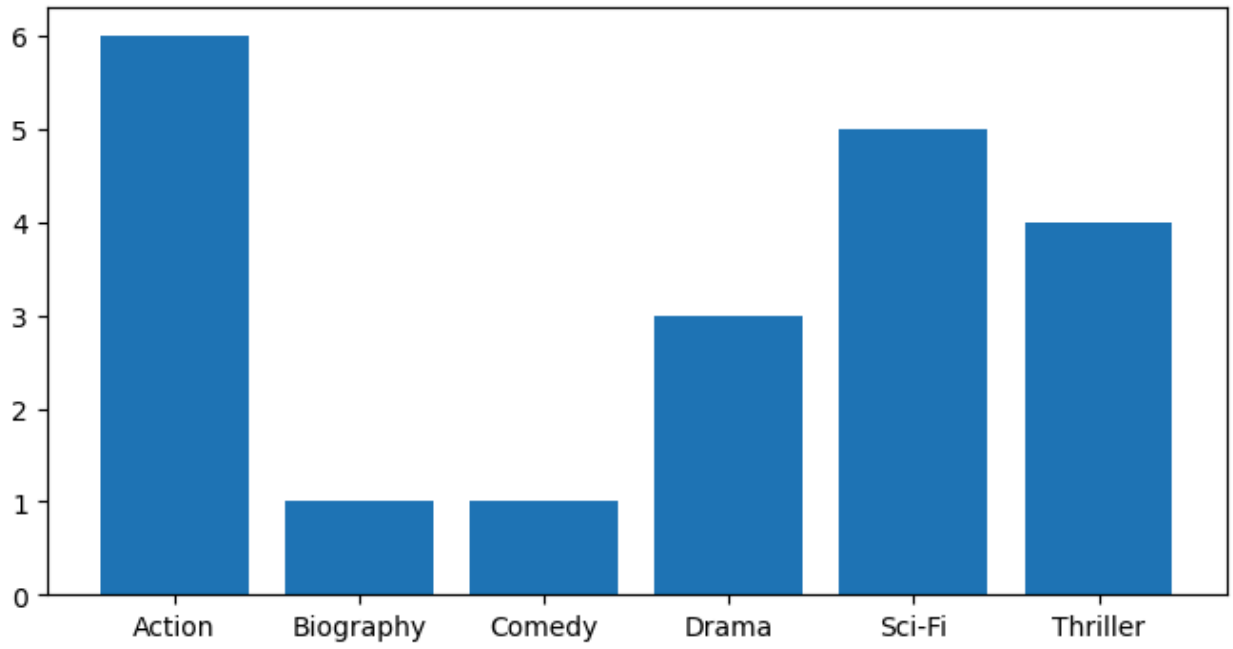
Top 10 movies by box office

In [5]:
```
top10 = df.sort_values(by='box_office_million',ascending=False).head(10)
plt.figure(figsize=(12,5))
plt.bar(top10['title'],top10['box_office_million'])
plt.show()
```



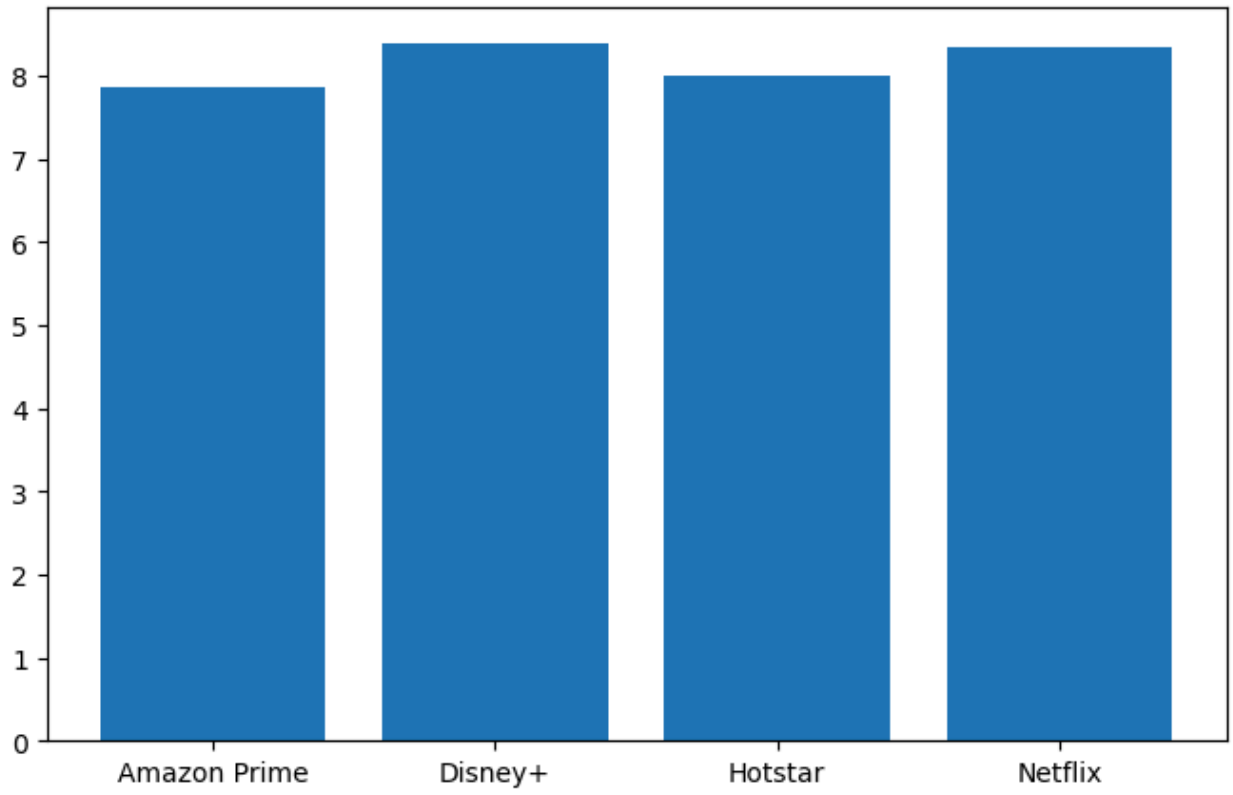
Number of movies per genre

In [6]:
```
gen_count = df.groupby('genre')['movie_id'].count()
plt.figure(figsize=(8,4))
plt.bar(gen_count.index,gen_count.values)
plt.show()
```



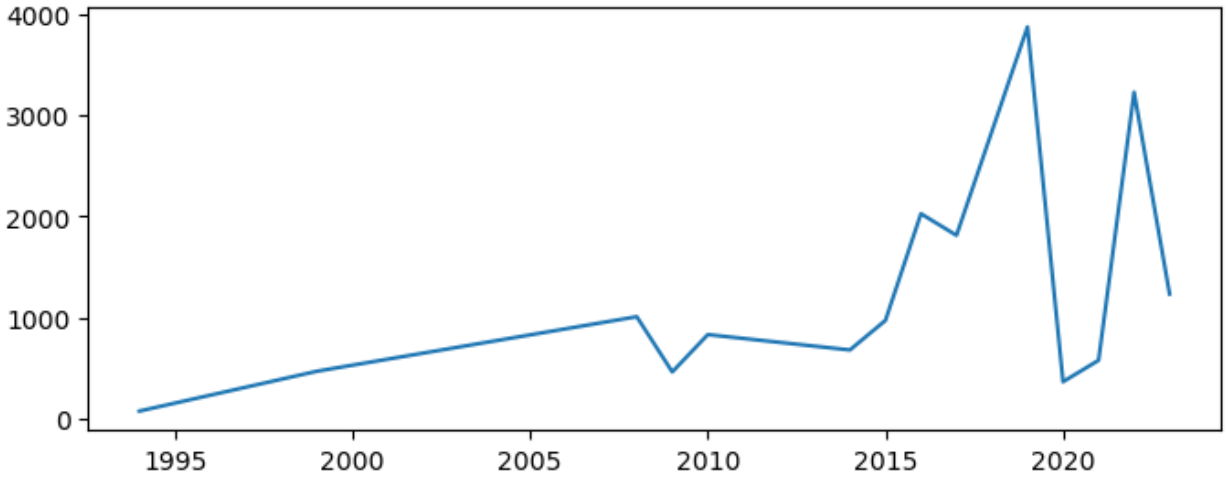
IMDb rating by platform

In [7]:
```
imd = df.groupby('platform')['imdb_rating'].mean().round(2)
plt.figure(figsize=(8,5))
plt.bar(imd.index,imd.values)
plt.show()
```



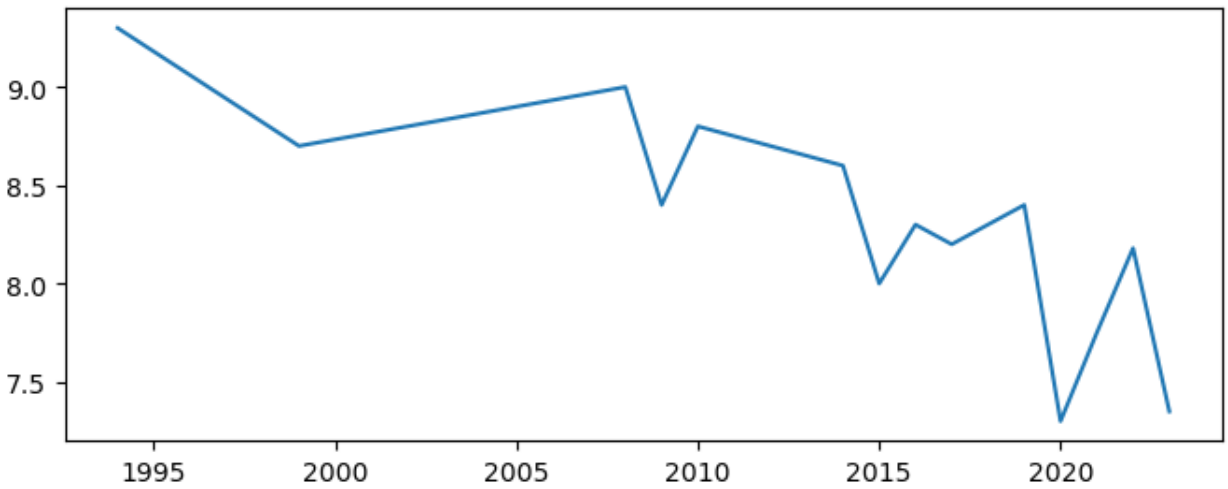
Box office collection trend over years

In [8]:
```
coll = df.groupby ('release_year')['box_office_million'].sum()
plt.figure(figsize=(8,3))
plt.plot(coll.index,coll.values)
plt.show()
```



Rating trend across years

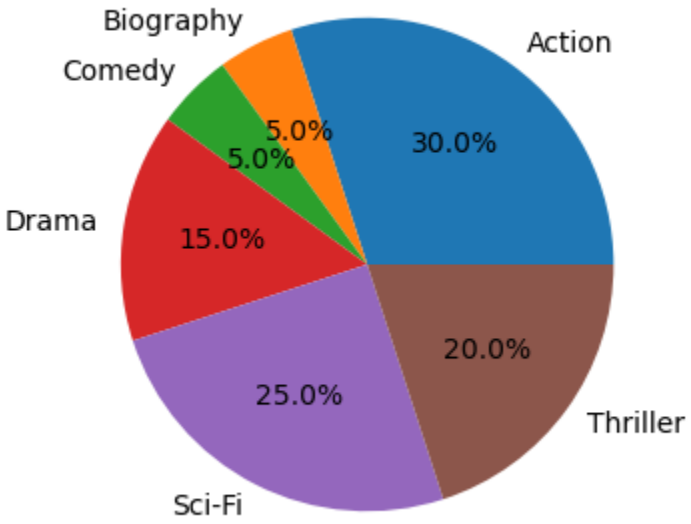
In [9]:
```
rating = df.groupby('release_year')['imdb_rating'].mean().round(2)
plt.figure(figsize=(8,3))
plt.plot(rating.index,rating.values)
plt.show()
```



Genre distribution

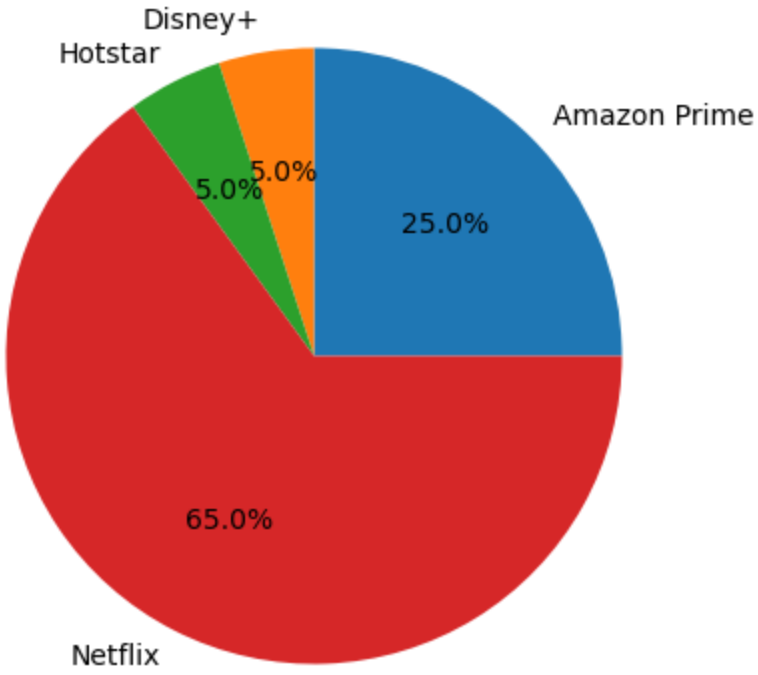
In [10]:
```
dis = df.groupby('genre')['movie_id'].count()
plt.figure(figsize=(4,4))
plt.pie(dis.values,labels = dis.index,autopct='%1.1f%%')

plt.show()
```



Platform distribution

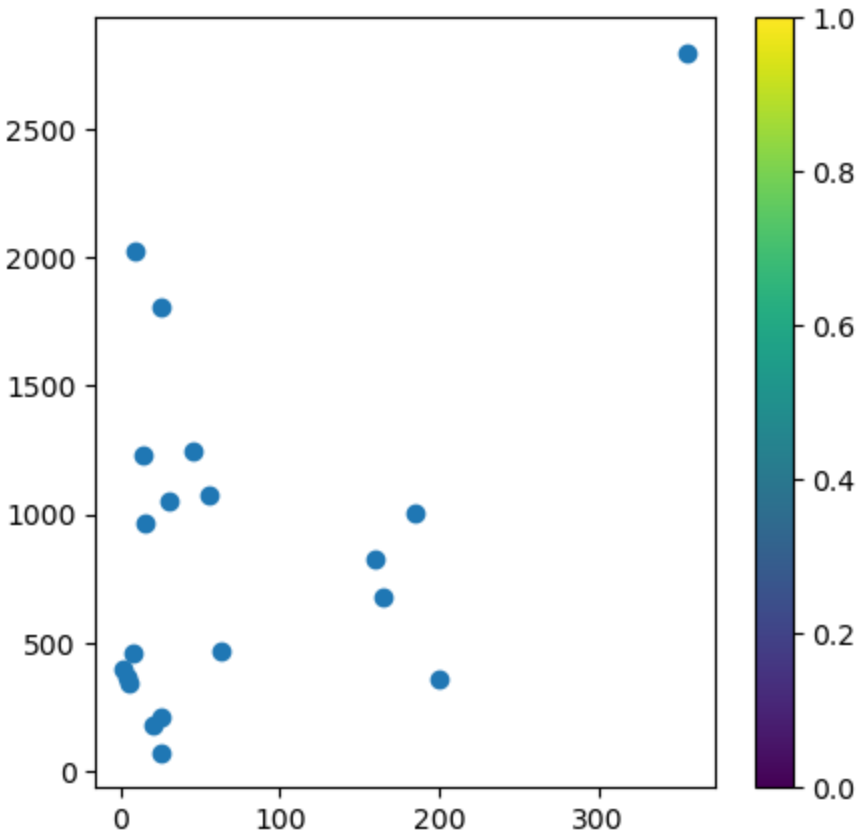
In [11]:
```
plat = df.groupby('platform')['movie_id'].count()
plt.figure(figsize=(5,5))
plt.pie(plat.values , labels=plat.index,autopct='%1.1f%%')
plt.show()
```



Budget vs Box Office

In [12]:
```
plt.figure(figsize=(5,5))
plt.scatter(df['budget_million'],df['box_office_million'],cmap='viridis')
plt.colorbar()
plt.show()
```

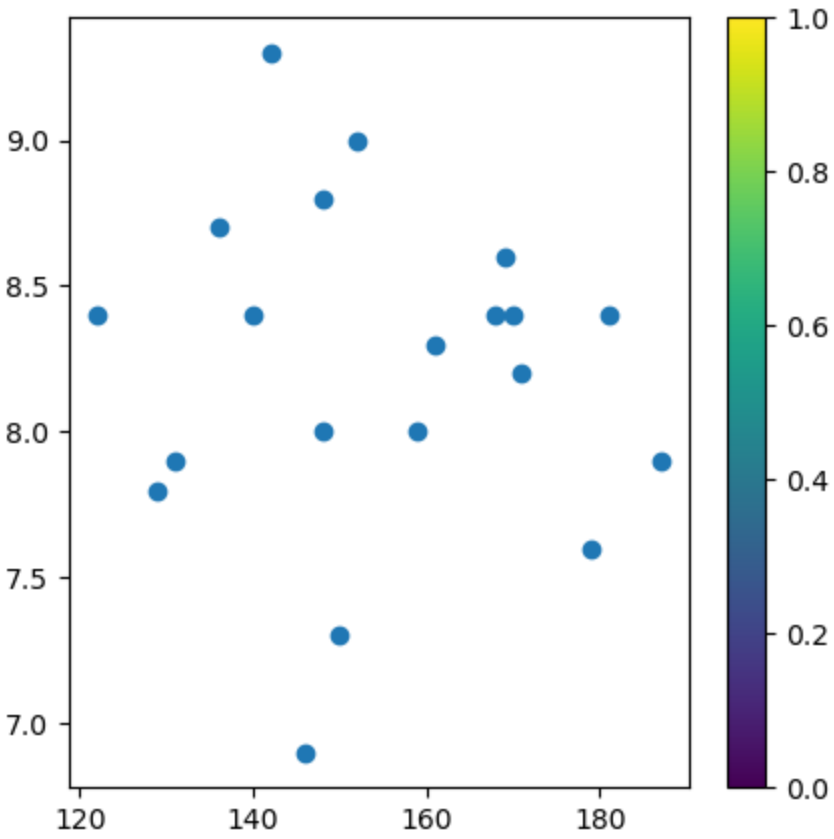
```
C:\Users\hp\AppData\Local\Temp\ipykernel_1296\2524496925.py:2: UserWarning: No
data for colormapping provided via 'c'. Parameters 'cmap' will be ignored
  plt.scatter(df['budget_million'],df['box_office_million'],cmap='viridis')
```



Duration vs IMDb Rating

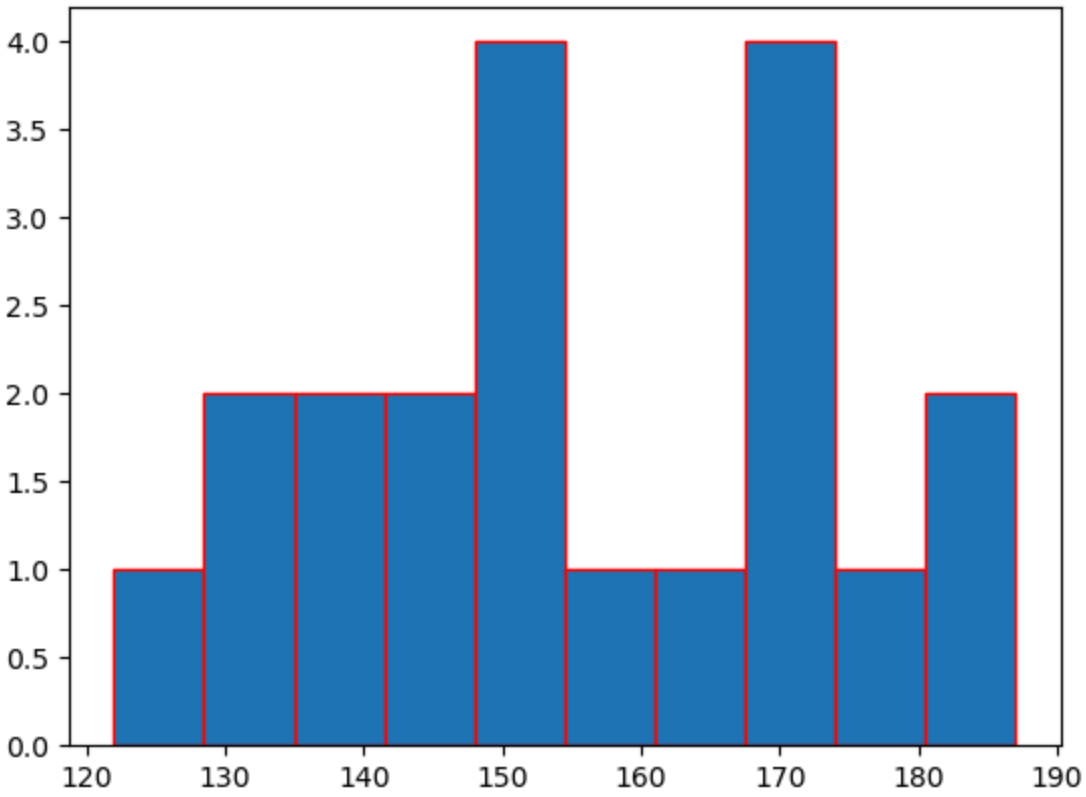
In [13]:
```
plt.figure(figsize=(5,5))
plt.scatter(df['duration_min'],df['imdb_rating'],cmap='viridis')
plt.colorbar()
plt.show()
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_1296\1978449638.py:2: UserWarning: No
data for colormapping provided via 'c'. Parameters 'cmap' will be ignored
  plt.scatter(df['duration_min'],df['imdb_rating'],cmap='viridis')
```



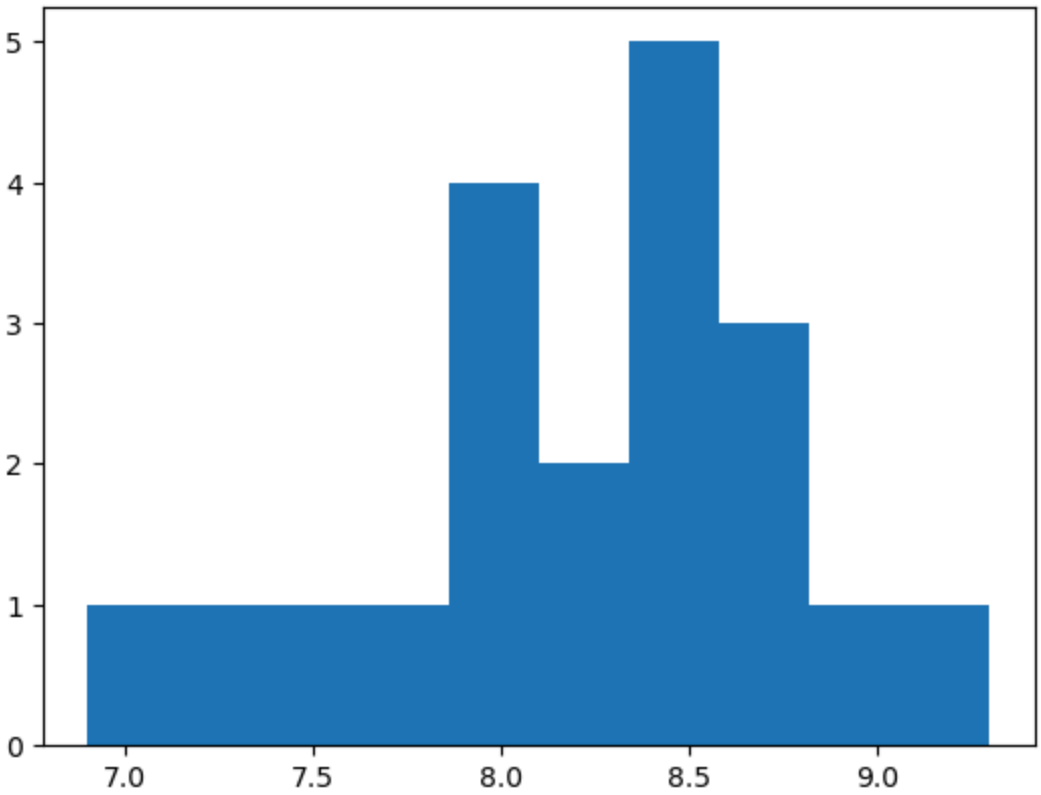
Movie duration distribution

In [14]:
```
plt.hist(df['duration_min'],edgecolor='r')
plt.show()
```



IMDb rating distribution

In [15]:
```
plt.hist(df['imdb_rating'])
plt.show()
```



In [ ]: