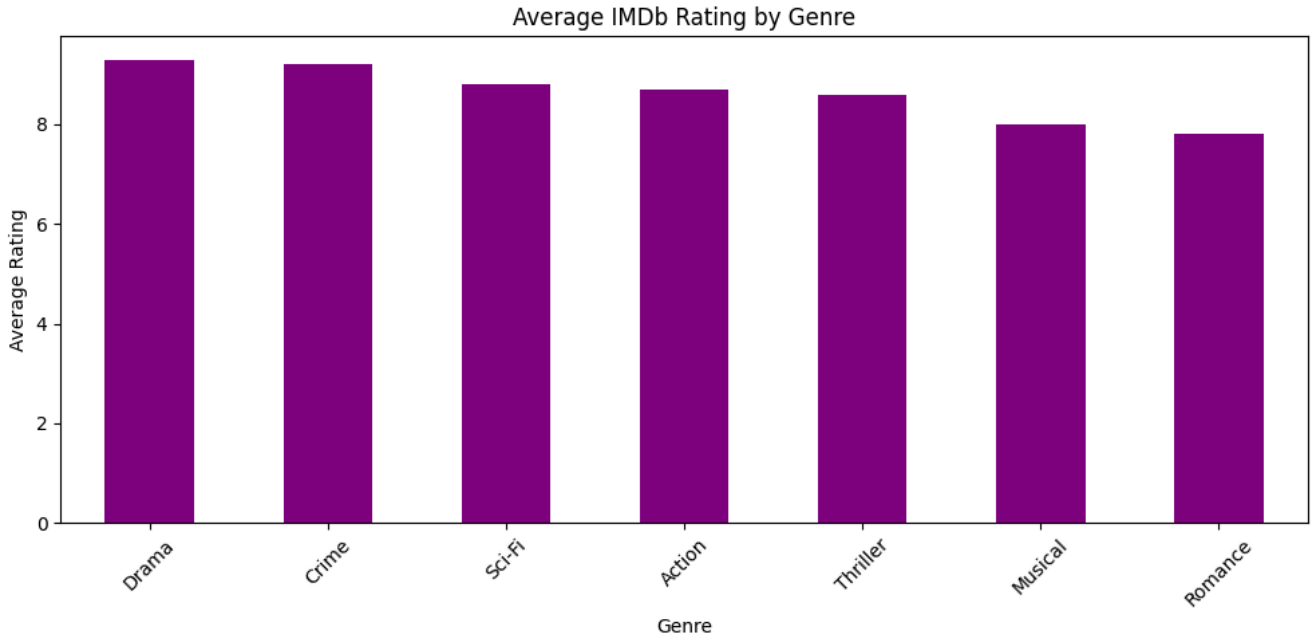
```
# Define the data as a multi-line string
csv_data = """Title,Genre,Director,Year,Runtime,Rating,Votes,Revenue (Millions),Metascore
Inception,Sci-Fi,Christopher Nolan,2010,148,8.8,2000000,825.5,74
Titanic,Romance,James Cameron,1997,195,7.8,1050000,2187.5,75
The Godfather,Crime,Francis Ford Coppola,1972,175,9.2,1600000,134.9,100
The Dark Knight,Action,Christopher Nolan,2008,152,9.0,2300000,1004.9,84
Avengers: Endgame,Action,Anthony Russo,2019,181,8.4,1100000,2797.8,78
La La Land,Musical,Damien Chazelle,2016,128,8.0,500000,446.1,93
Parasite,Thriller,Bong Joon-ho,2019,132,8.6,600000,266.9,96
The Shawshank Redemption,Drama,Frank Darabont,1994,142,9.3,2500000,58.3,80
"""

# Specify the filename
filename = "movies.csv"

# Open the file in write mode ('w')
# This will create the file if it doesn't exist, or overwrite it if it does
with open(filename, 'w') as f:
    # Write the data to the file
    f.write(csv_data)

print(f"Data successfully saved to {filename}")
```

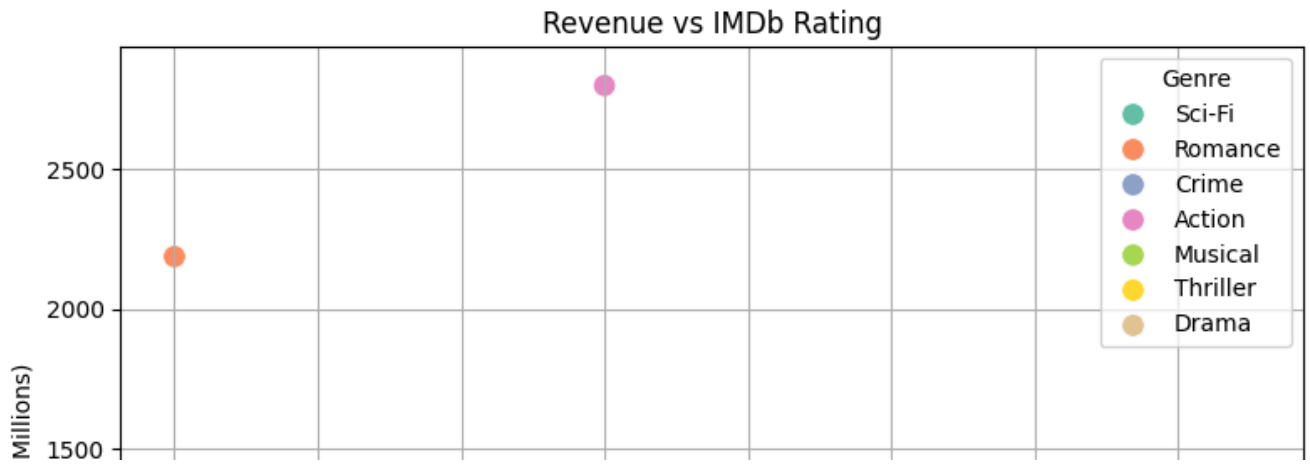
```
Data successfully saved to movies.csv
```

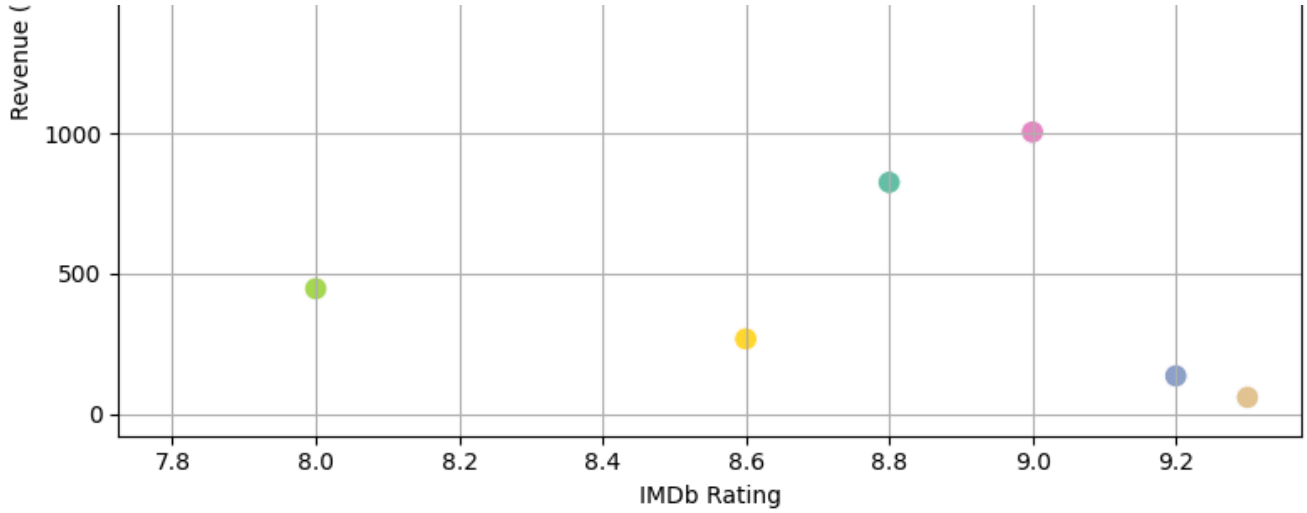
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
# Load dataset
df = pd.read_csv("movies.csv")

# Show basic info
print("📊 Dataset Info:")
print(df.info())
print("\n🔍 First 5 Rows:")
print(df.head())

# ------------------------------
# 1. Average IMDb Rating by Genre
# ------------------------------
plt.figure(figsize=(10, 5))
genre_ratings = df.groupby('Genre')['Rating'].mean().sort_values(ascending=False)
genre_ratings.plot(kind='bar', color='purple')
plt.title("Average IMDb Rating by Genre")
plt.xlabel("Genre")
plt.ylabel("Average Rating")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# ------------------------------
# 2. Revenue vs Rating Scatter Plot
# ------------------------------
plt.figure(figsize=(8, 6))
sns.scatterplot(data=df, x='Rating', y='Revenue (Millions)', hue='Genre', palette='Set2', s=100)
plt.title("Revenue vs IMDb Rating")
plt.xlabel("IMDb Rating")
plt.ylabel("Revenue (Millions)")
plt.grid(True)
plt.tight_layout()
plt.show()
```

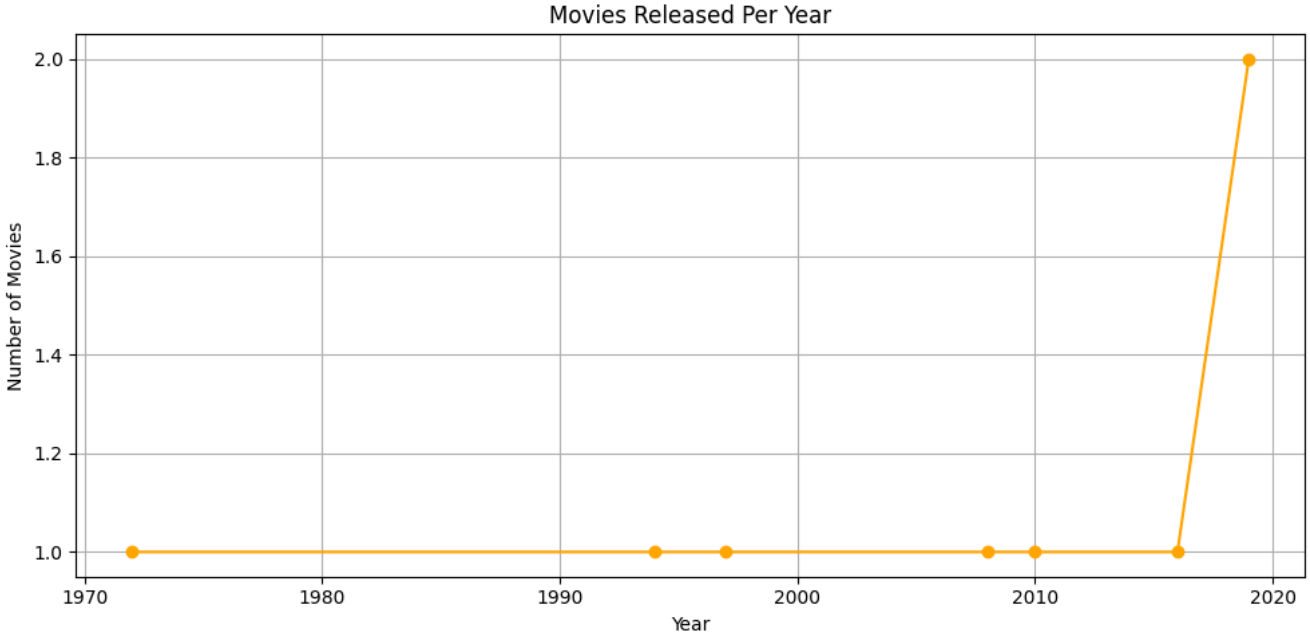
```
# ----------------------------
# 3. Number of Movies Released Per Year
# ----------------------------
plt.figure(figsize=(10, 5))
df['Year'].value_counts().sort_index().plot(kind='line', marker='o', color='orange')
plt.title("Movies Released Per Year")
plt.xlabel("Year")
plt.ylabel("Number of Movies")
plt.grid(True)
plt.tight_layout()
plt.show()

# ----------------------------
# 4. Correlation Heatmap
# ----------------------------
plt.figure(figsize=(8, 5))
correlation_matrix = df[['Rating', 'Votes', 'Revenue (Millions)', 'Runtime', 'Metascore']].corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Feature Correlation Heatmap")
plt.tight_layout()
plt.show()

# ----------------------------
# 5. Top Directors by High Ratings
# ----------------------------
high_rated = df[df['Rating'] >= 8.0]
top_directors = high_rated['Director'].value_counts().head(5)
plt.figure(figsize=(8, 4))
top_directors.plot(kind='barh', color='teal')
plt.title("Top 5 Directors with IMDb Rating ≥ 8.0")
plt.xlabel("Number of High-Rated Movies")
plt.gca().invert_yaxis()
plt.tight_layout()
plt.show()


#visualization 6
# 6. Interactive Bubble Plot: Budget vs Revenue
fig = px.scatter(df,
                 x="Rating",
                 y="Revenue (Millions)",
                 size="Votes",
                 color="Genre",
                 hover_name="Title",
                 title="Interactive: Rating vs Revenue (bubble by Votes)",
                 template="plotly_dark")
fig.show()
```

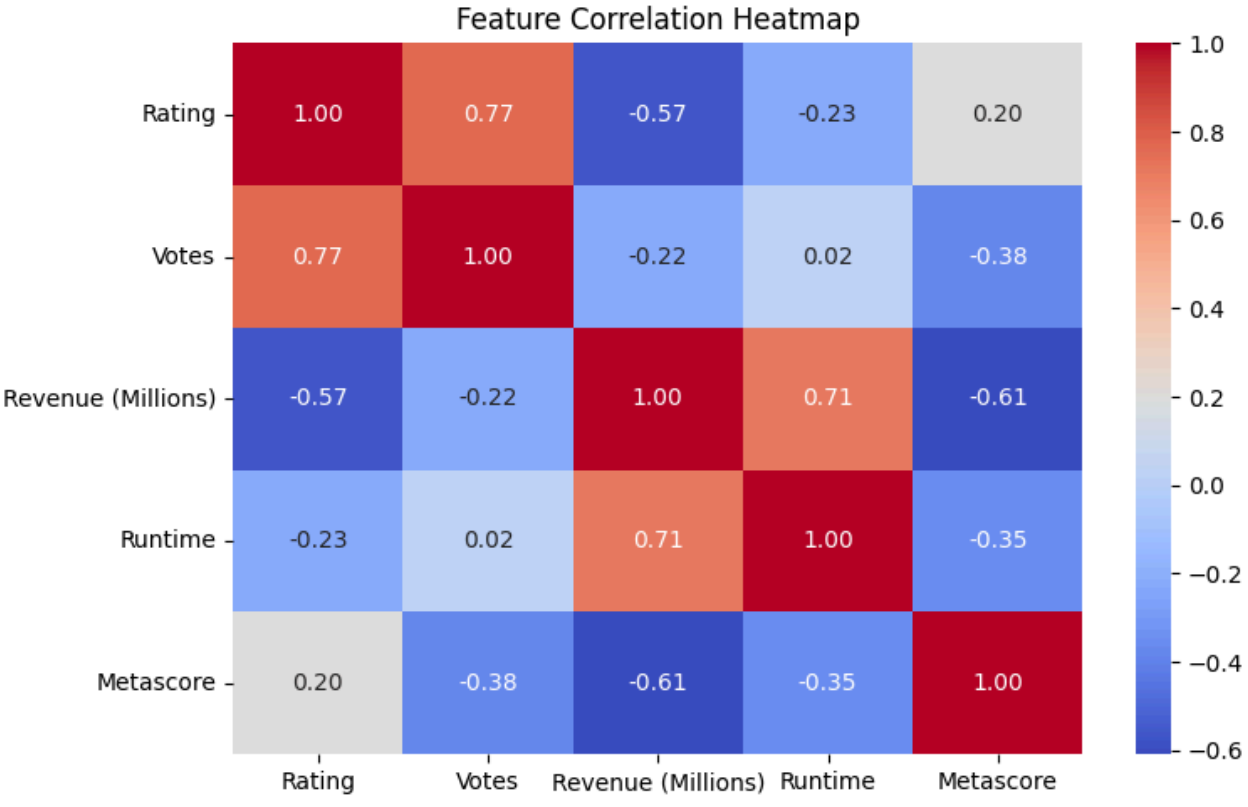
```
📊 Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 9 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Title               8 non-null      object
 1   Genre               8 non-null      object
 2   Director            8 non-null      object
 3   Year                8 non-null      int64
 4   Runtime             8 non-null      int64
 5   Rating              8 non-null      float64
 6   Votes               8 non-null      int64
 7   Revenue (Millions)  8 non-null      float64
 8   Metascore           8 non-null      int64
dtypes: float64(2), int64(4), object(3)
memory usage: 708.0+ bytes
None

🔍 First 5 Rows:
              Title    Genre             Director  Year  Runtime  Rating  \
0         Inception   Sci-Fi     Christopher Nolan  2010      148     8.8
1           Titanic  Romance         James Cameron  1997      195     7.8
2      The Godfather    Crime  Francis Ford Coppola  1972      175     9.2
3    The Dark Knight   Action     Christopher Nolan  2008      152     9.0
4  Avengers: Endgame   Action         Anthony Russo  2019      181     8.4

     Votes  Revenue (Millions)  Metascore
0  2000000               825.5         74
1  1050000              2187.5         75
2  1600000               134.9        100
3  2300000              1004.9         84
4  1100000              2797.8         78
```



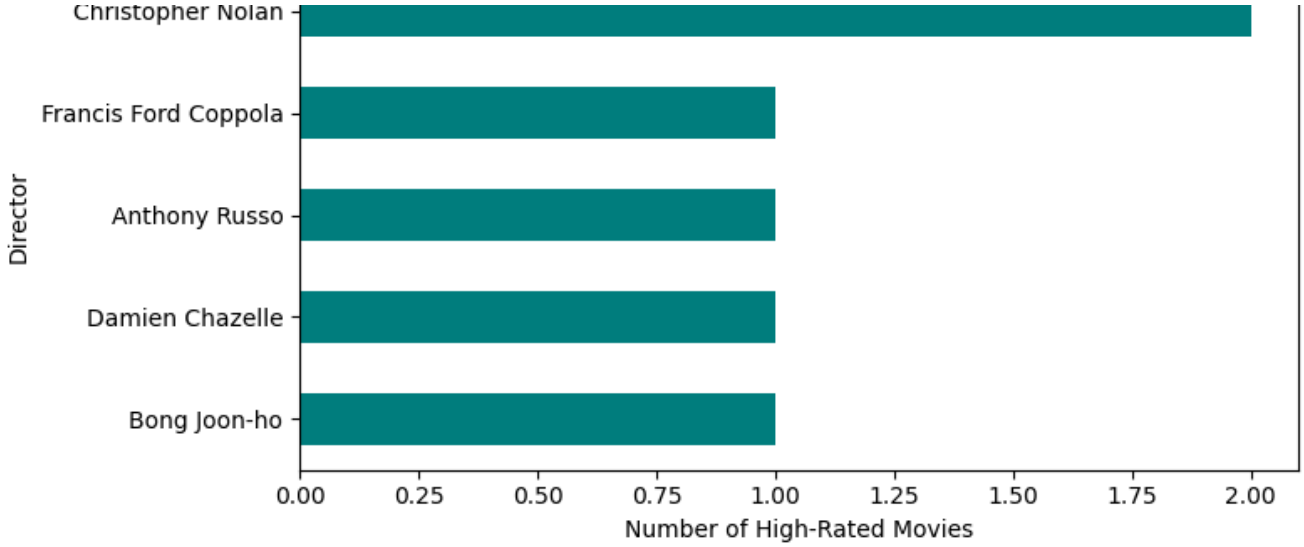



Revenue (
1000
500
0
7.8
8.0
8.2
8.4
8.6
8.8
9.0
9.2
IMDb Rating


Movies Released Per Year
Number of Movies
2.0
1.8
1.6
1.4
1.2
1.0
1970
1980
1990
2000
2010
2020
Year


Feature Correlation Heatmap
Rating
Votes
Revenue (Millions)
Runtime
Metascore
1.00 0.77 -0.57 -0.23 0.20
0.77 1.00 -0.22 0.02 -0.38
-0.57 -0.22 1.00 0.71 -0.61
-0.23 0.02 0.71 1.00 -0.35
0.20 -0.38 -0.61 -0.35 1.00
Rating
Votes
Revenue (Millions)
Runtime
Metascore
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


Top 5 Directors with IMDb Rating ≥ 8.0



Interactive: Rating vs Revenue (bubble by Votes)