```python
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("Netflix.csv")
# Clean data: drop rows with missing values in important columns
df = df.dropna(subset=["type", "title", "duration", "genres", "description"])
# Count occurrences of each type (e.g., Movie, TV Show)
type_count = df['type'].value_counts()
# Plot the counts as a bar chart
plt.figure(figsize=(6, 4))
plt.bar(type_count.index, type_count.values, color=["red","green"])
plt.title("Count of Content Types on Netflix")
plt.xlabel("Type")
plt.ylabel("Count")
plt.tight_layout()
plt.savefig("netflix.png",dpi=300,bbox_inches='tight')
plt.show()
```
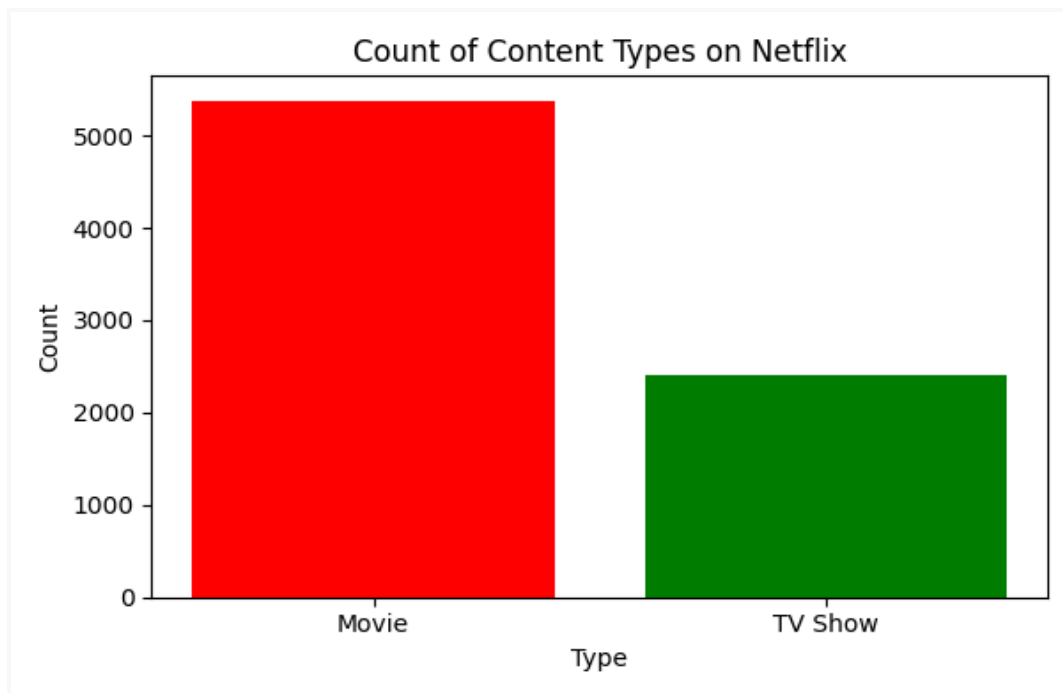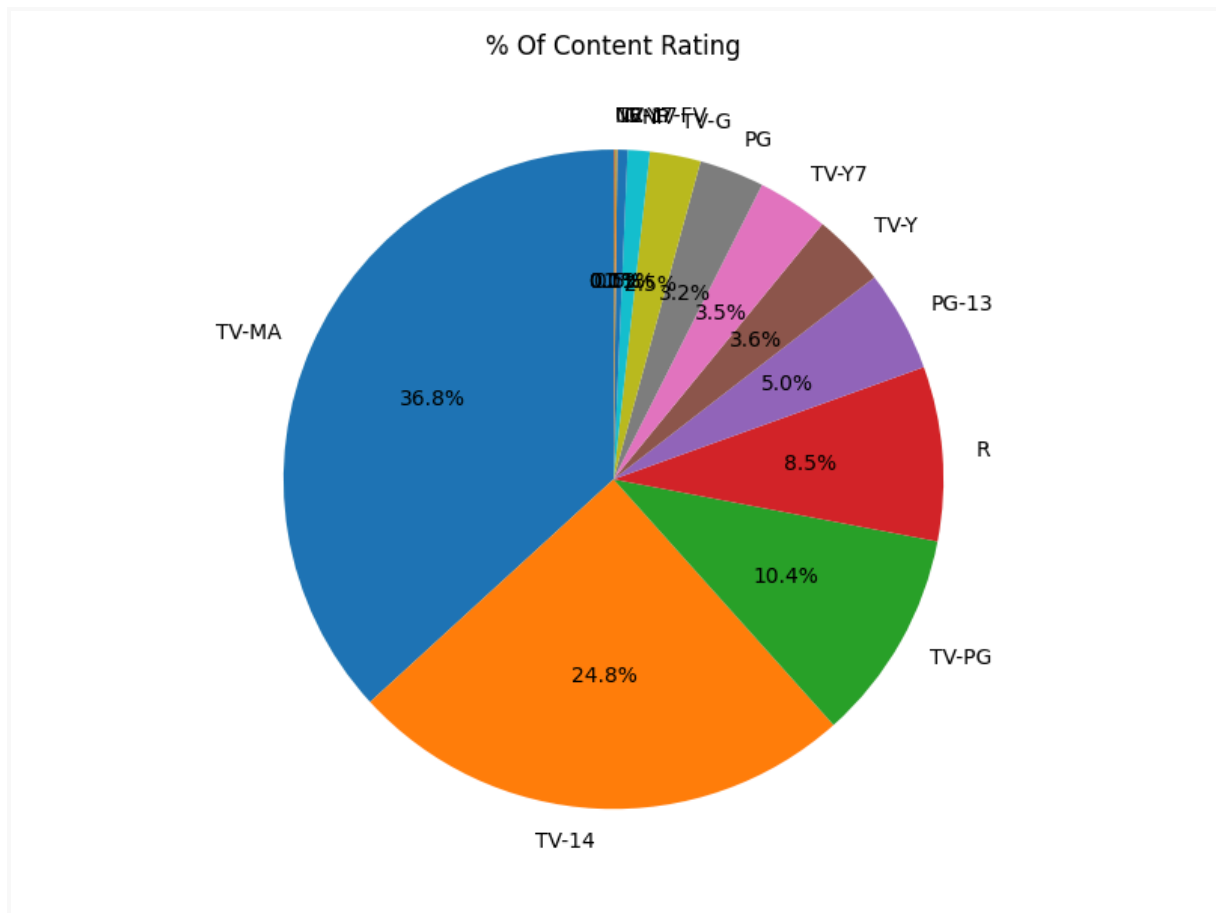


```python
# Assuming df is already loaded and cleaned
rating_count = df['rating'].value_counts()
plt.figure(figsize=(8, 6))
plt.pie(rating_count, labels=rating_count.index, autopct='%1.1f%%', startangle=90)
plt.title("% Of Content Rating")
plt.tight_layout()
plt.savefig("netu.png")
plt.show()
```

% Of Content Rating


```python
# Filter only movies
movie_df = df[df['type'] == "Movie"].copy()

# Remove ' min' and convert to integer safely
movie_df = movie_df.dropna(subset=['duration'])  # Optional: drop rows with missing duration
movie_df['duration_int'] = movie_df['duration'].astype(str).str.replace(' min', '').astype(int)

# Plotting the histogram
plt.figure(figsize=(8, 6))
plt.hist(movie_df['duration_int'], bins=30, color='purple', edgecolor='black')
plt.title('Distribution of Movie Duration')
plt.xlabel('Duration (minutes)')
plt.ylabel('Number of Movies')
plt.tight_layout()
plt.savefig('movie_duration_histogram.png')
plt.show()
```
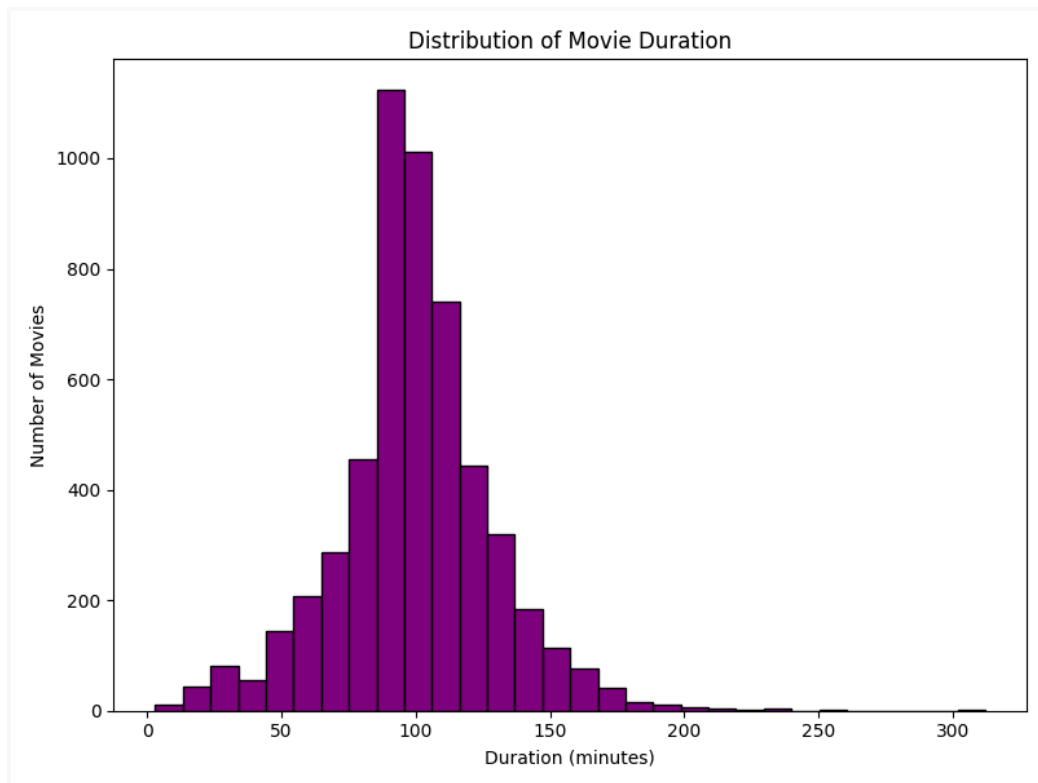
Distribution of Movie Duration

```
release_counts = df['release_year'].value_counts().sort_index()
plt.figure(figsize=(10,6))
plt.scatter(release_counts.index, release_counts.values, color='red')
plt.title('Release Year VS Number of Shows')
plt.xlabel('Release Year')
plt.ylabel('Number of Shows')
plt.tight_layout()
plt.savefig('release_year_Scatter.png')
plt.show()
```


Release Year VS Number of Shows
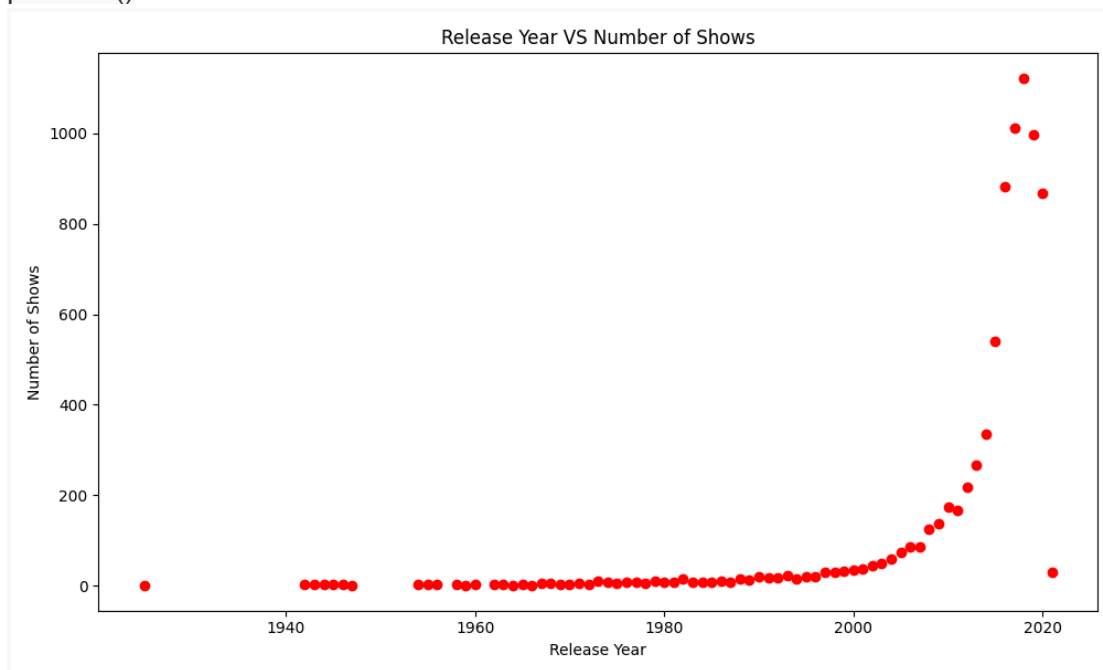
```python
country_count = df['country'].value_counts().head(10)
plt.figure(figsize=(8,6))
plt.barh(country_count.index, country_count.values, color="teal")
plt.title('Top 10 Countries by Number of Shows')
plt.xlabel('Number of Shows')
plt.ylabel('Country')
plt.tight_layout()
plt.savefig('top10_countries.png')
plt.show()
```
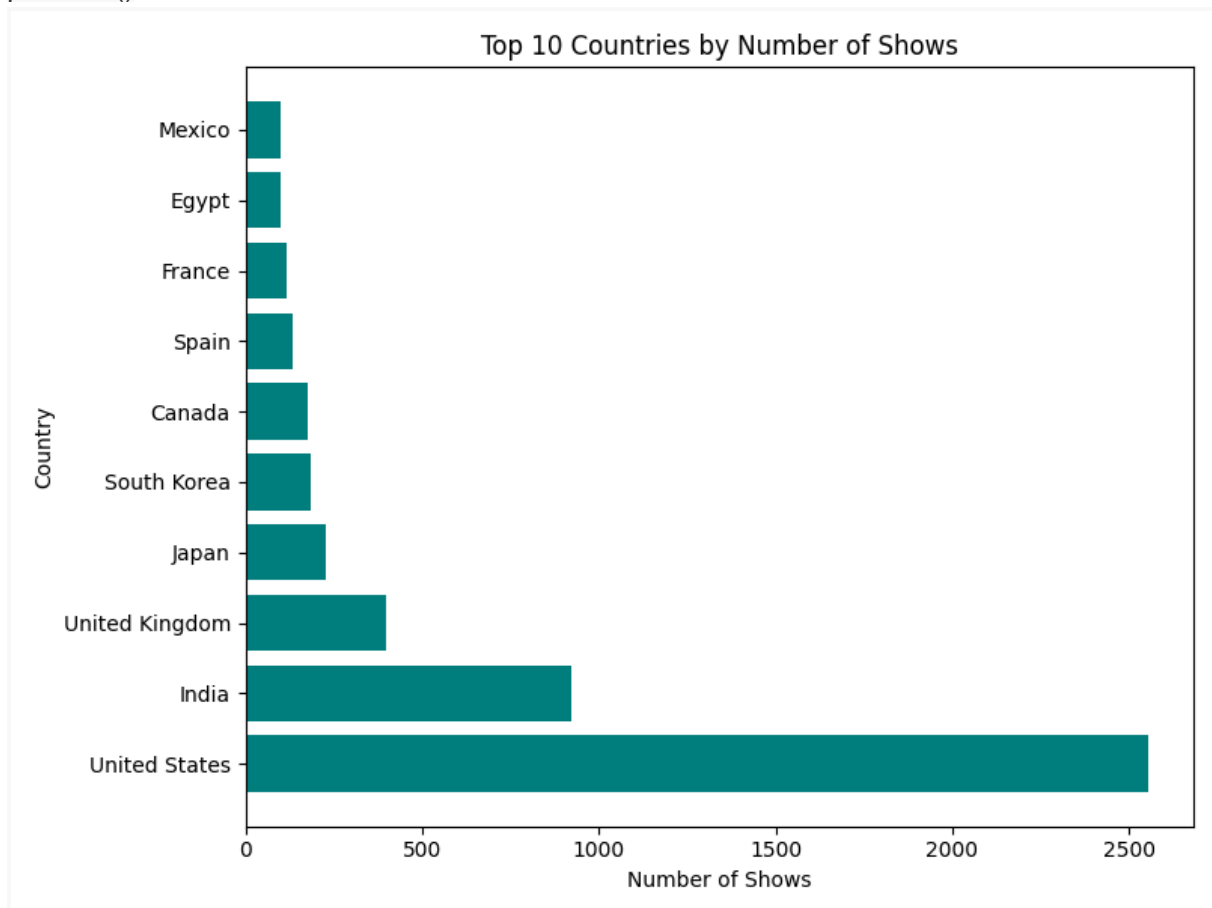


```python
content_by_year = df.groupby(['release_year', 'type']).size().unstack().fillna(0)

fig, ax = plt.subplots(1, 2, figsize=(12, 5))

# first subplot: Movies
ax[0].plot(content_by_year.index, content_by_year['Movie'], color='blue')
ax[0].set_title('Movies Released Per Year')
ax[0].set_xlabel('Year')
ax[0].set_ylabel('Number of Movies')

# second subplot: TV Shows
ax[1].plot(content_by_year.index, content_by_year['TV Show'], color='orange')
ax[1].set_title('TV Shows Released Per Year')
ax[1].set_xlabel('Year')
```

```
ax[1].set_ylabel('Number of Movies')

fig.suptitle('Comparison of Movies and TV Shows Released Over Years')

plt.tight_layout()
plt.savefig('movie_tv_year_comparison.png')
plt.show()
```