net

May 17, 2025

[75]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud
```

[13]:
```
print(data.isnull().sum())
```

```
show_id         0
type            0
title           0
director        0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
dtype: int64
```

[17]:
```
data.drop_duplicates(inplace=True)
```

[29]:
```
data.dropna(subset=['director','title','country'], inplace=True)
```

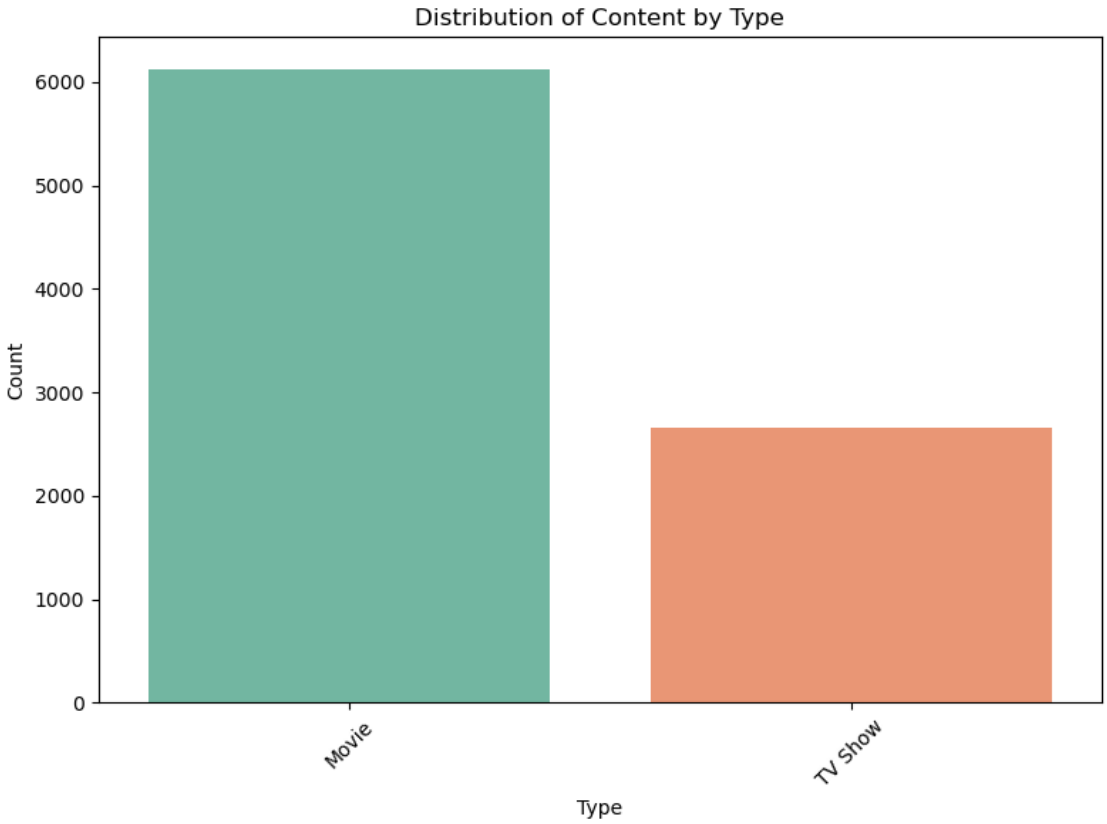
[31]:
```
data['date_added'] = pd.to_datetime(data['date_added'])
```

[33]:
```
print(data.dtypes)
```

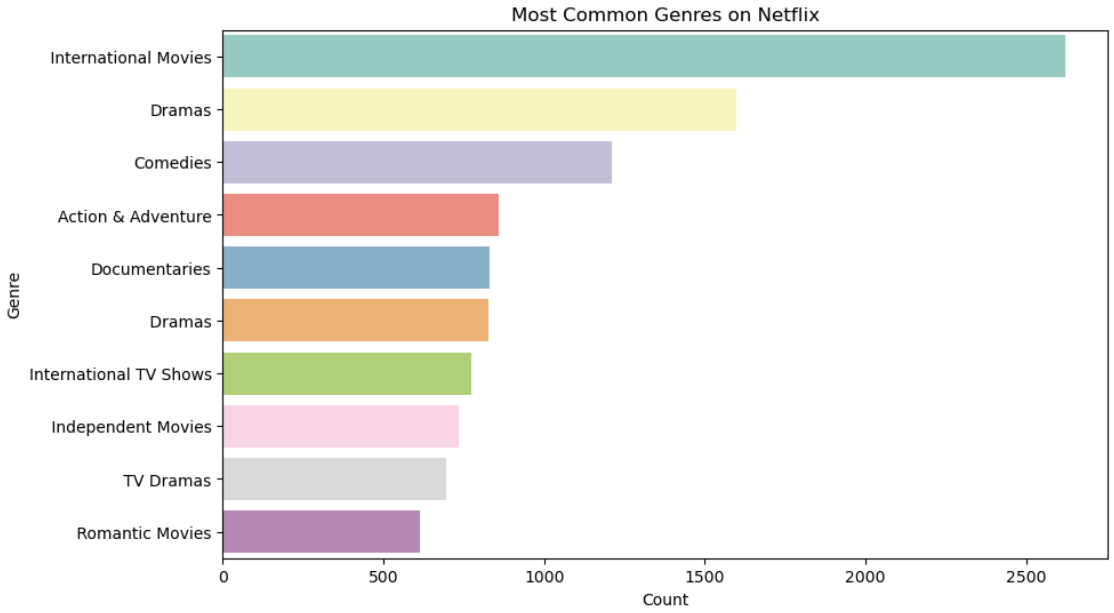
```
show_id                 object
type                    object
title                   object
director                object
country                 object
date_added      datetime64[ns]
release_year             int64
rating                  object
duration                object
```

```
listed_in               object
dtype: object
```

[83]:
```
# Create a DataFrame for plotting
type_counts = data['type'].value_counts().reset_index()
type_counts.columns = ['Type', 'Count']

# Plot
plt.figure(figsize=(8, 6))
sns.barplot(
    data=type_counts,
    x='Type',
    y='Count',
    hue='Type',           # Set hue equal to x
    palette='Set2',
    legend=False          # Suppress the redundant legend
)

# Add titles and labels
plt.title('Distribution of Content by Type')
plt.xlabel('Type')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

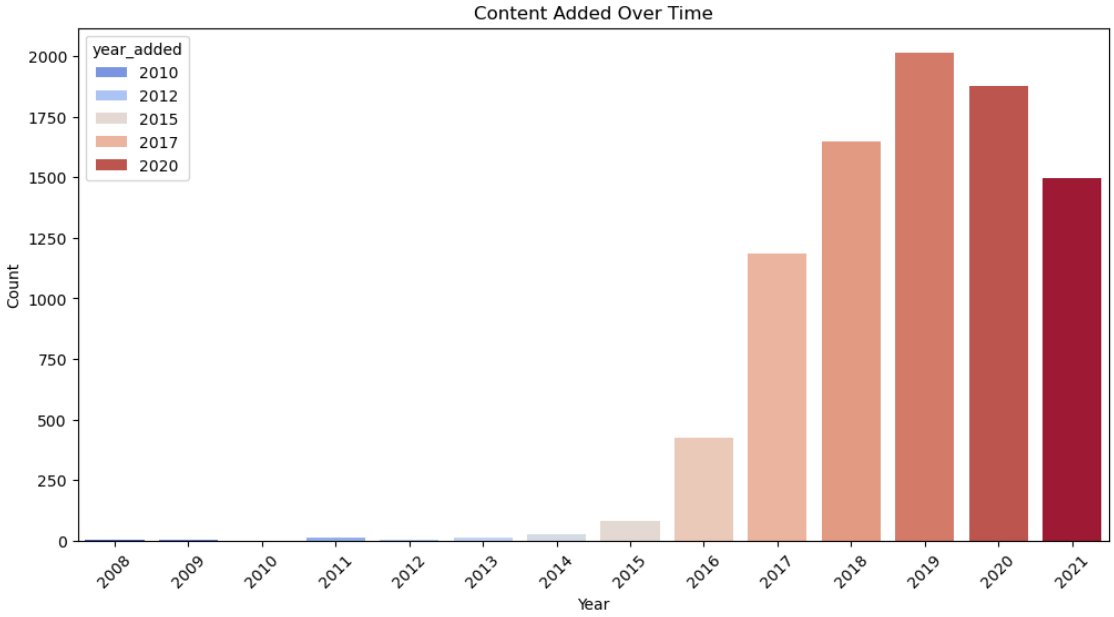


[97]:
```
data['genres'] = data['listed_in'].apply(lambda x: x.split(','))
all_genres = sum(data['genres'], [])
genre_counts = pd.Series(all_genres).value_counts().head(10)


plt.figure(figsize=(10, 6))
sns.barplot(x=genre_counts.values, y=genre_counts.index,hue=genre_counts.
 ↪index,palette='Set3')
plt.title('Most Common Genres on Netflix')
plt.xlabel('Count')
plt.ylabel('Genre')
plt.show()
```



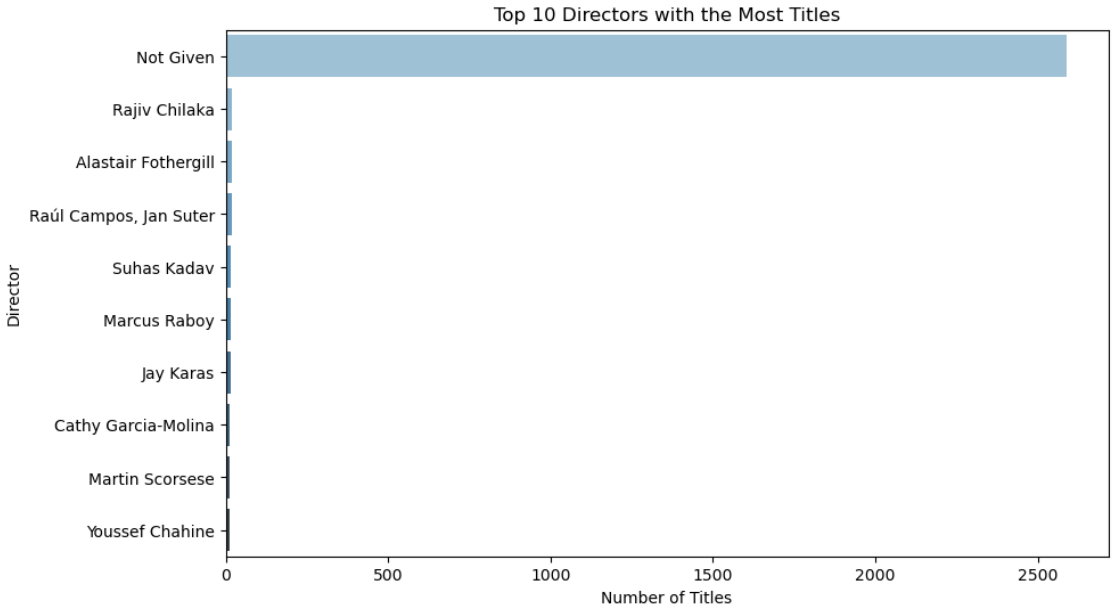
[53]:
```
data['year_added'] = data['date_added'].dt.year
data['month_added'] = data['date_added'].dt.month
```

[99]:
```
plt.figure(figsize=(12, 6))
sns.countplot(x='year_added', data=data, palette='coolwarm', hue='year_added')
plt.title('Content Added Over Time')
plt.xlabel('Year')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```

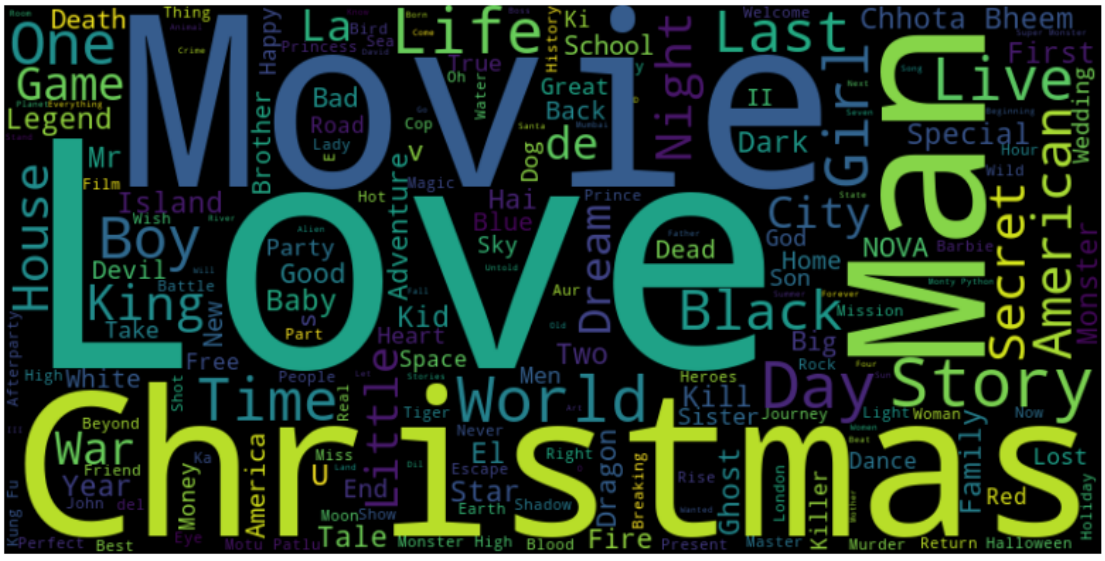


[63]:
```
top_directors = data['director'].value_counts().head(10)
```

[101]:
```
# Count titles by director
top_directors = data['director'].value_counts().head(10)
# Plot top directors
plt.figure(figsize=(10, 6))
sns.barplot(x=top_directors.values, y=top_directors.index,palette='Blues_d',␣
 ↪hue=top_directors.index)
plt.title('Top 10 Directors with the Most Titles')
plt.xlabel('Number of Titles')
plt.ylabel('Director')
plt.show()
```



[77]:
```
movie_titles = data[data['type'] =='Movie']['title']
wordcloud = WordCloud(width=800, height=400,background_color='black').
  ↪generate(' '.join(movie_titles))
# Plot word cloud
plt.figure(figsize=(10, 6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```



[ ]: