```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```
from google.colab import files
uploaded = files.upload()
```

Browse... netflix_titles.csv

netflix_titles.csv(application/vnd.ms-excel) - 3399671 bytes, last modified: n/a - 100% done

```
Saving netflix_titles.csv to netflix_titles.csv
```

```
import os
os.listdir()
```

```
['.config', 'netflix_titles.csv', 'sample_data']
```

```
df = pd.read_csv("netflix_titles.csv")
df.head()
```

	show_id	type	title	director	cast	country	date_added	releas
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy	NaN	September 24, 2021	

Next steps: Generate code with df New interactive sheet

```
# Check for missing values
df.isnull().sum()

# Drop duplicates
df.drop_duplicates(inplace=True)

# Fill missing values (example)
df['country'].fillna("Unknown", inplace=True)
df['director'].fillna("Not Specified", inplace=True)
df['cast'].fillna("Not Specified", inplace=True)
df['rating'].fillna("Not Rated", inplace=True)
```

```
/tmp/ipython-input-1524088280.py:8: FutureWarning: A value is trying to be se
The behavior will change in pandas 3.0. This inplace method will never work b

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.


  df['country'].fillna("Unknown", inplace=True)
/tmp/ipython-input-1524088280.py:9: FutureWarning: A value is trying to be se
The behavior will change in pandas 3.0. This inplace method will never work b

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.


  df['director'].fillna("Not Specified", inplace=True)
/tmp/ipython-input-1524088280.py:10: FutureWarning: A value is trying to be s
The behavior will change in pandas 3.0. This inplace method will never work b

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.


  df['cast'].fillna("Not Specified", inplace=True)
/tmp/ipython-input-1524088280.py:11: FutureWarning: A value is trying to be s
The behavior will change in pandas 3.0. This inplace method will never work b

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.


  df['rating'].fillna("Not Rated", inplace=True)
```

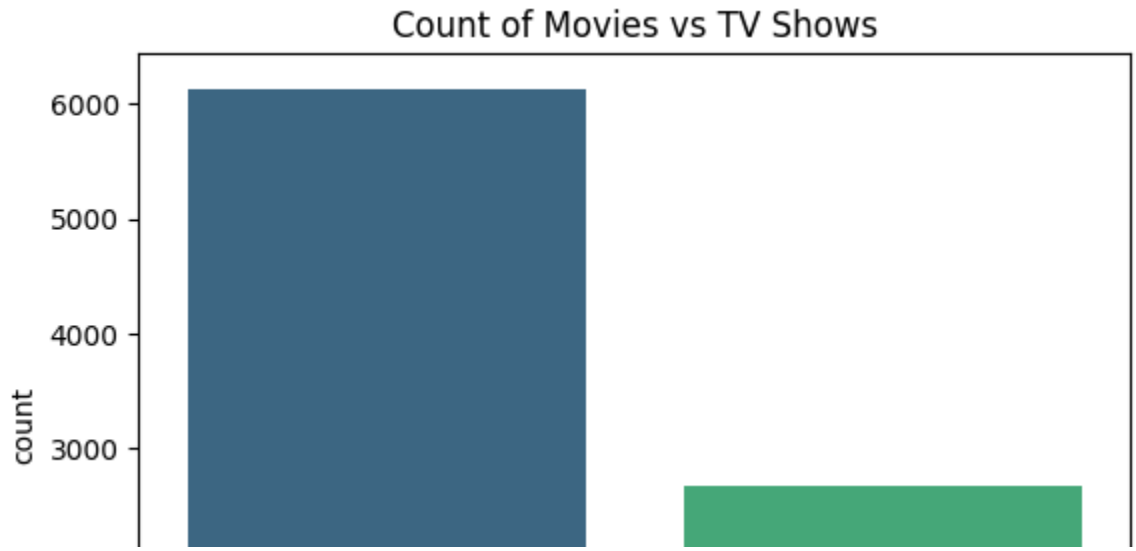
```
print(df.info())
print(df.describe())
print(df['type'].value_counts())
```

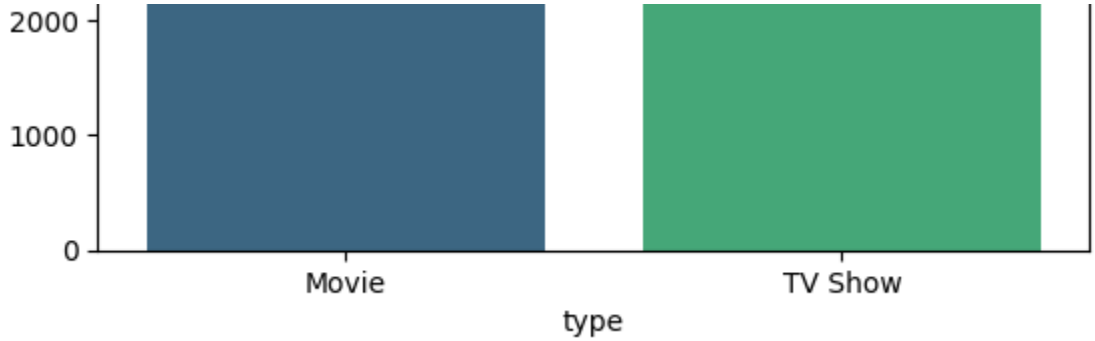
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       8807 non-null   object
```

```
 1   type          8807 non-null   object
 2   title         8807 non-null   object
 3   director      8807 non-null   object
 4   cast          8807 non-null   object
 5   country       8807 non-null   object
 6   date_added    8797 non-null   object
 7   release_year  8807 non-null   int64
 8   rating        8807 non-null   object
 9   duration      8804 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
None
       release_year
count   8807.000000
mean    2014.180198
std        8.819312
min     1925.000000
25%     2013.000000
50%     2017.000000
75%     2019.000000
max     2021.000000
type
Movie      6131
TV Show    2676
Name: count, dtype: int64
```

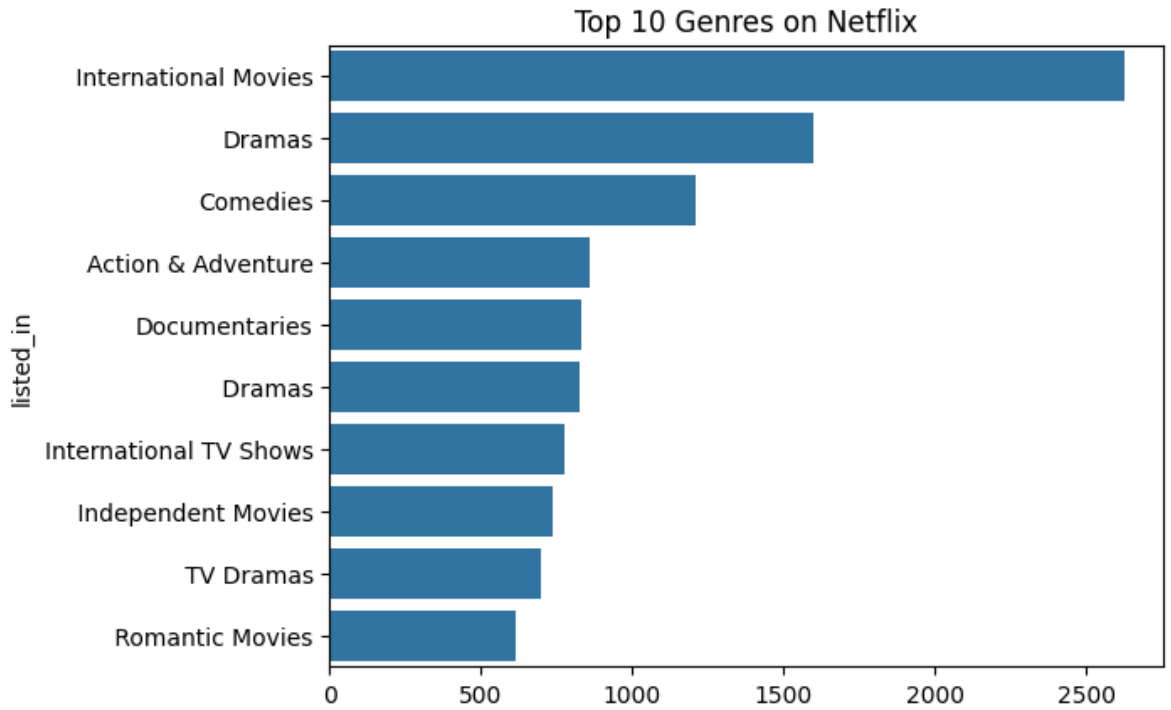
```
sns.countplot(x='type', data=df, palette='viridis')
plt.title('Count of Movies vs TV Shows')
plt.show()
```

```
/tmp/ipython-input-1992240860.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed i

  sns.countplot(x='type', data=df, palette='viridis')
```





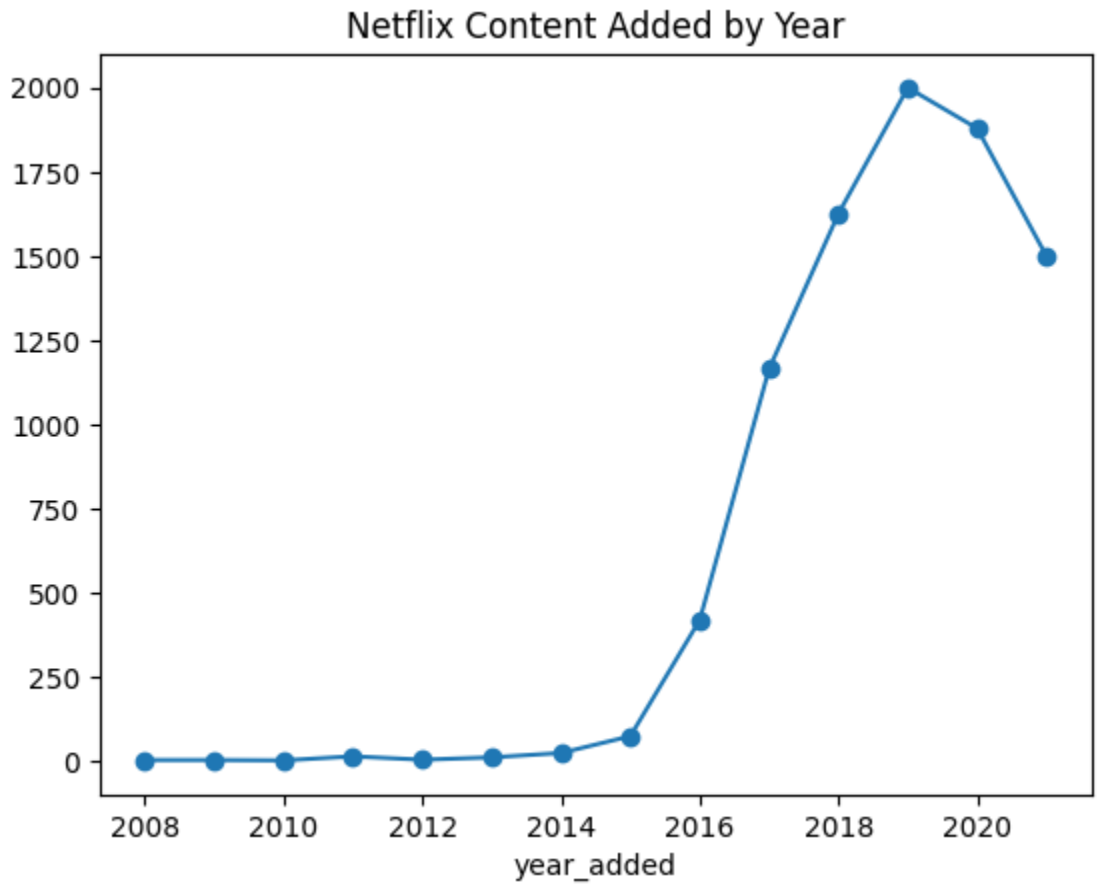

```
genres = df['listed_in'].str.split(',').explode().value_counts().head(10)
sns.barplot(y=genres.index, x=genres.values)
plt.title('Top 10 Genres on Netflix')
plt.show()
```



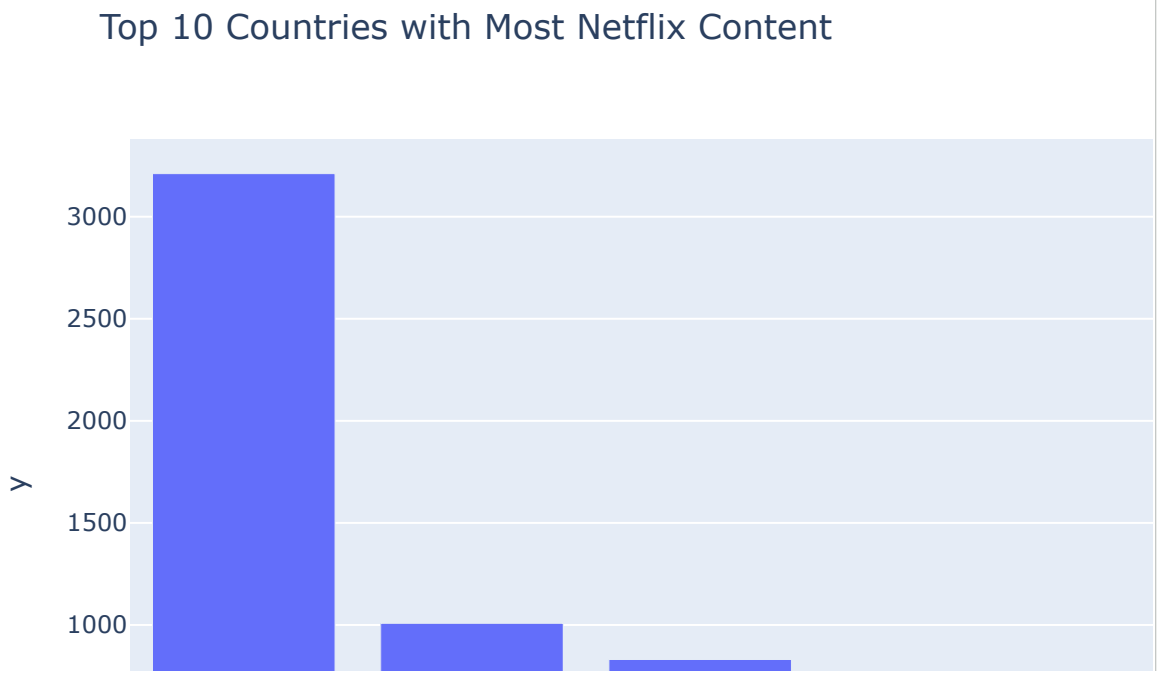

```
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce', infer_dat

df['year_added'] = df['date_added'].dt.year
content_by_year = df['year_added'].value_counts().sort_index()
content_by_year.plot(kind='line', marker='o', title='Netflix Content Added by
plt.show()
```

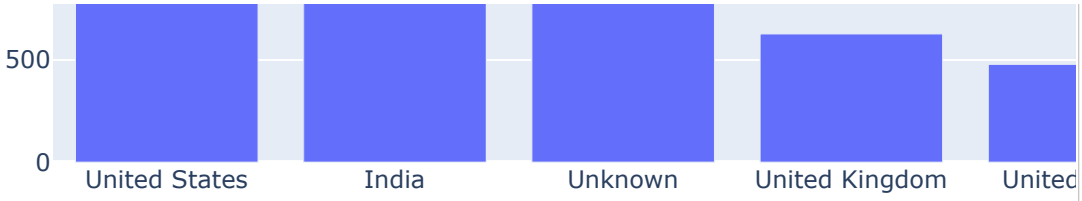
```
/tmp/ipython-input-4226973630.py:1: UserWarning: The argument 'infer_datetime
```

```
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce', inter_
```




```
countries = df['country'].str.split(',').explode().value_counts().head(10)
px.bar(x=countries.index, y=countries.values, title='Top 10 Countries with Mos
```






```
from wordcloud import WordCloud

text = ' '.join(df['listed_in'].dropna())
wordcloud = WordCloud(width=800, height=400, background_color='black').generat

plt.figure(figsize=(10,6))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Popular Genres on Netflix')
plt.show()
```