netflixdataexploration

October 23, 2024

NETFLIX : DATA EXPLORATION AND VISUALIZATION

[7]:
```
import pandas as pd

df = pd.read_csv('C:/Users/shrad/Downloads/NETFLIX.csv')

df.head()
```

[7]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm…
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act…	
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV …	

```
1  After crossing paths at a party, a Cape Town t…
2  To protect his family from a powerful drug lor…
3  Feuds, flirtations and toilet talk go down amo…
4  In a city of coaching centers known to train I…
```

[9]:
```
df.isnull().sum()
```

[9]:
```
show_id            0
type               0
title              0
director        2634
cast             825
country          831
date_added        10
release_year       0
rating             4
duration           3
listed_in          0
description        0
dtype: int64
```

[11]:
```
df.dtypes
```

[11]:
```
show_id         object
type            object
title           object
director        object
cast            object
country         object
date_added      object
release_year     int64
rating          object
duration        object
listed_in       object
description     object
dtype: object
```

[13]:
```
df.describe()
```

[13]:
```
       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
```

[15]:
```
df['type'].value_counts()
```

[15]:
```
type
Movie      6131
TV Show    2676
Name: count, dtype: int64
```

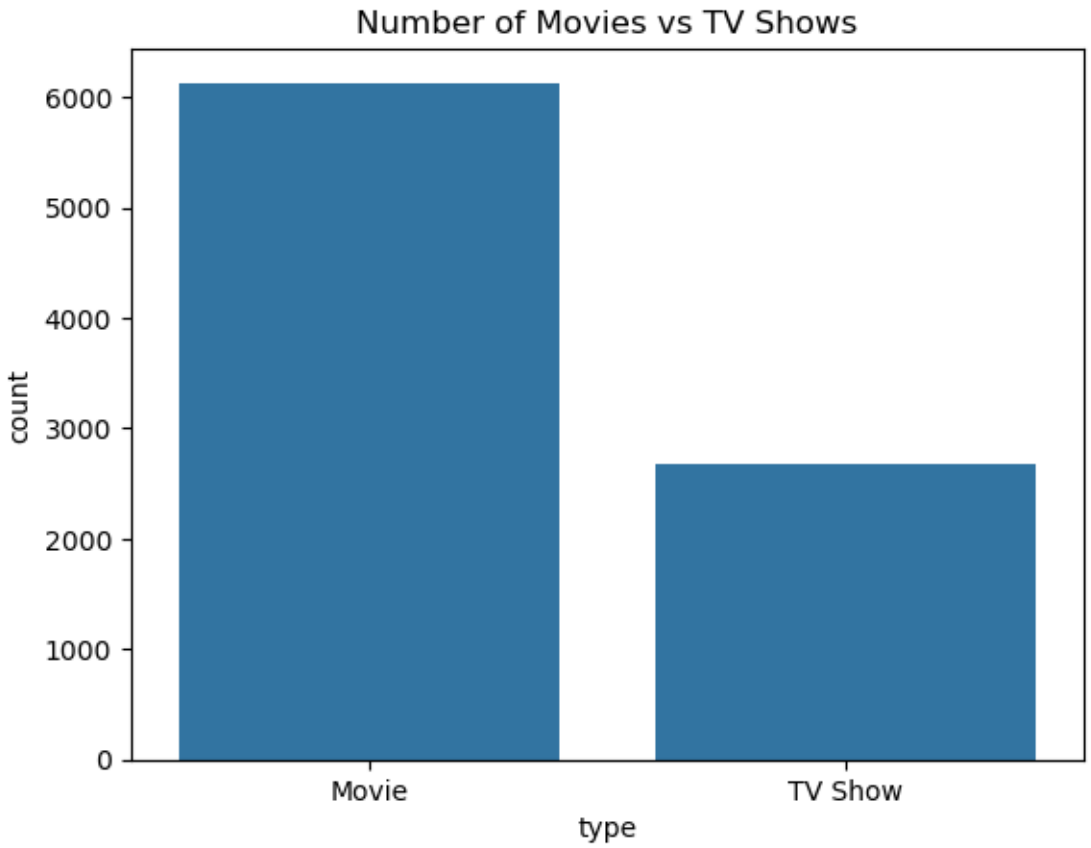
[17]:
```
df['country'].nunique(), df['listed_in'].nunique(), df['director'].nunique()
```

[17]:
```
(748, 514, 4528)
```

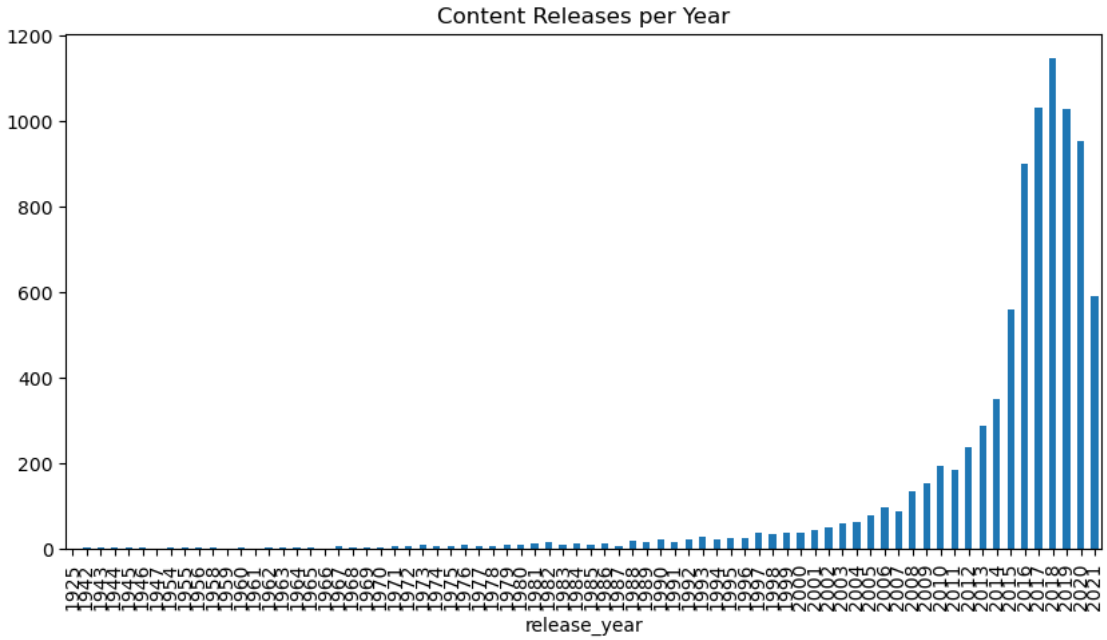
[19]:
```
df['release_year'].value_counts().sort_index()
```

[19]:
```
release_year
1925       1
1942       2
1943       3
1944       3
1945       4
          …
2017    1032
2018    1147
2019    1030
2020     953
2021     592
Name: count, Length: 74, dtype: int64
```

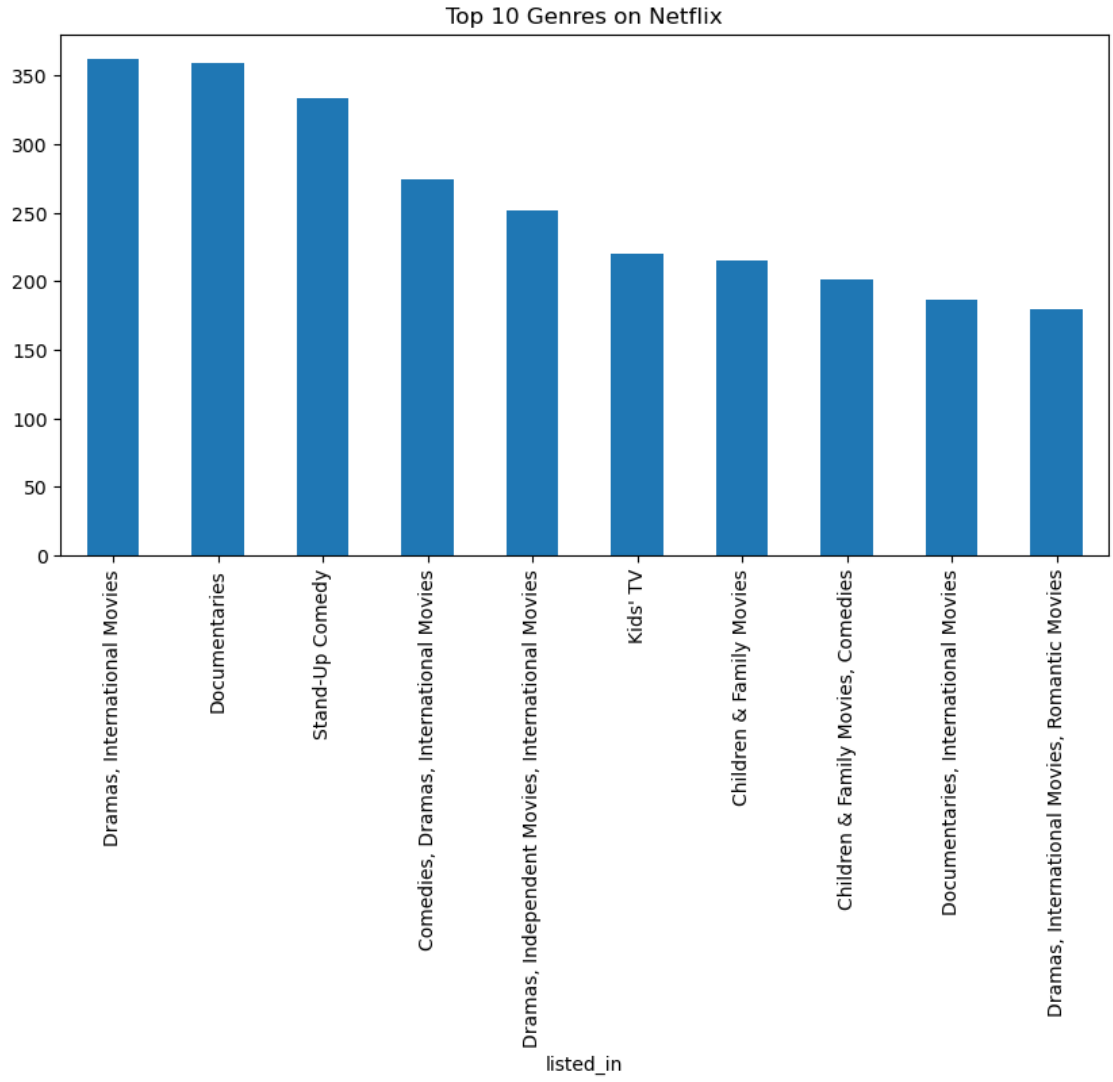
[21]:
```
import seaborn as sns
import matplotlib.pyplot as plt

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



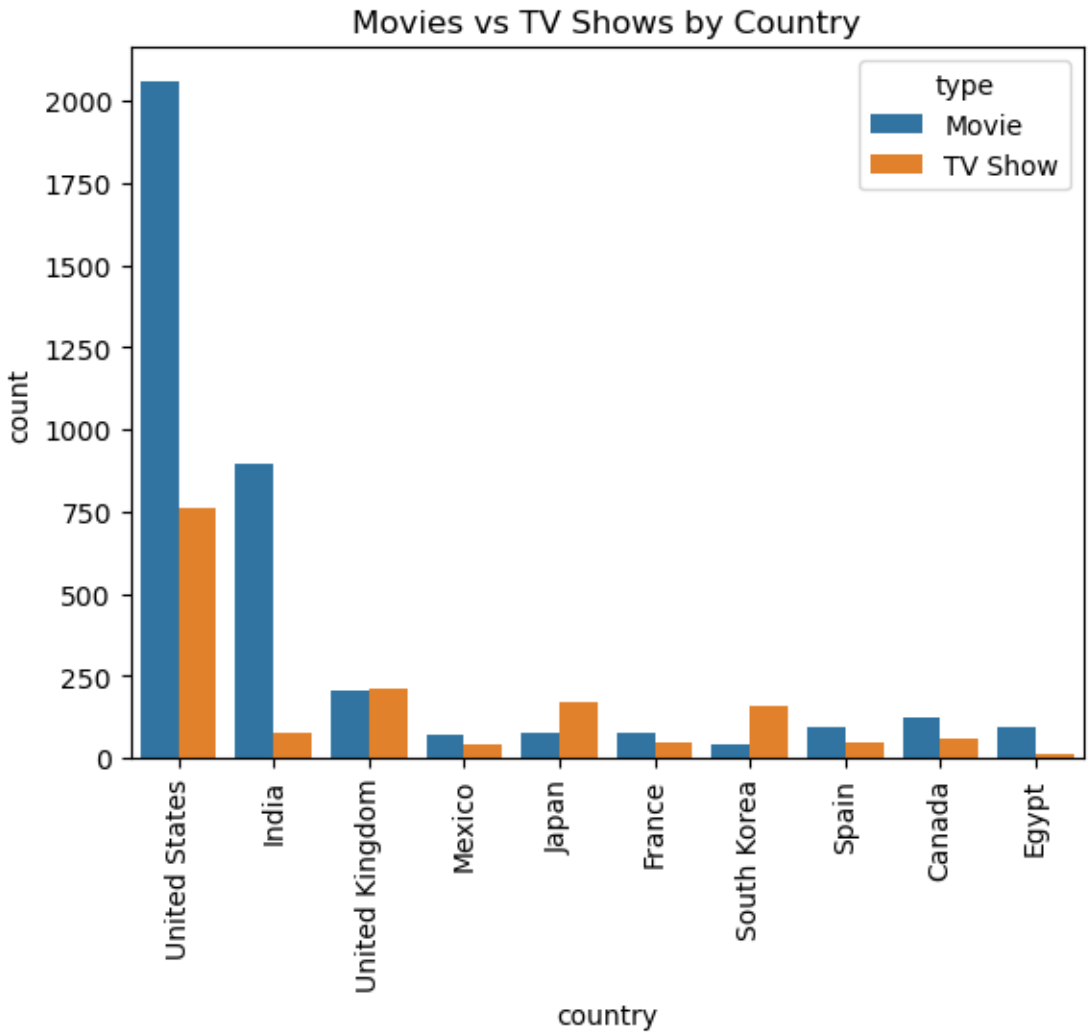
[23]:
```
df['release_year'].value_counts().sort_index().plot(kind='bar', figsize=(10,5))
plt.title('Content Releases per Year')
plt.show()
```



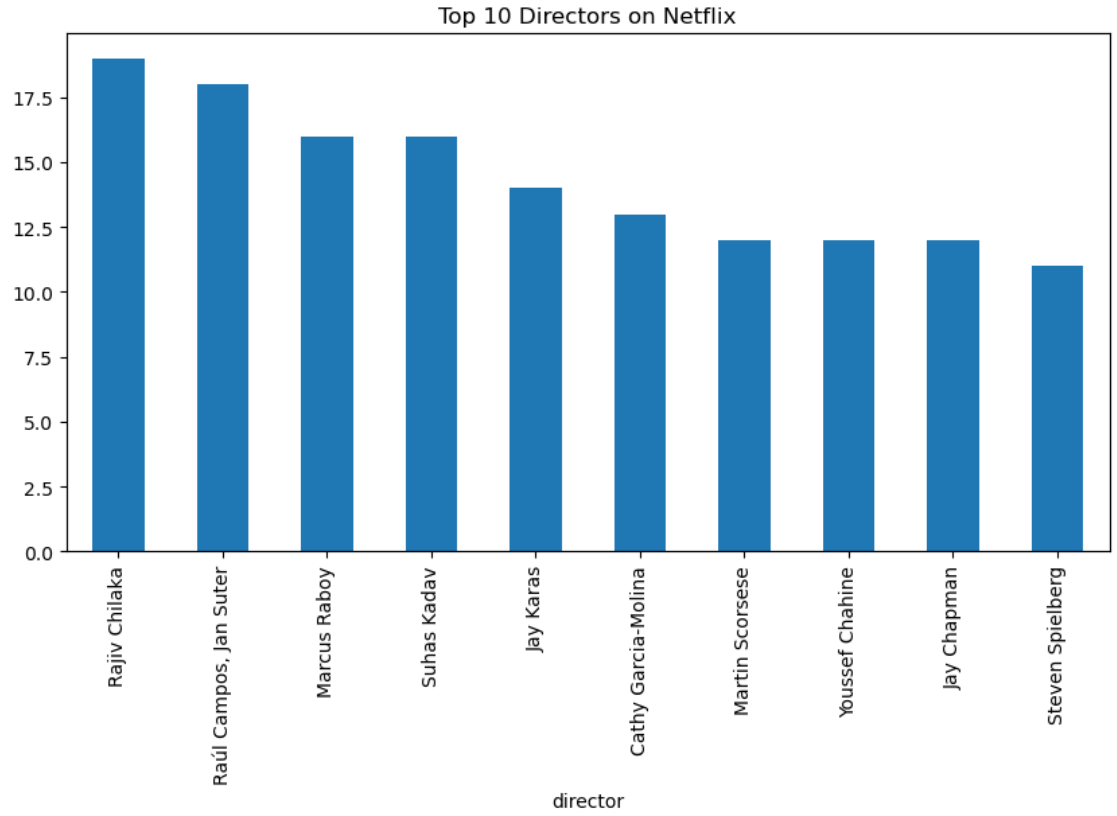
[26]:
```
df['listed_in'].value_counts().head(10).plot(kind='bar', figsize=(10,5))
plt.title('Top 10 Genres on Netflix')
plt.show()
```



[28]:
```
top_countries = df['country'].value_counts().head(10)

sns.countplot(data=df[df['country'].isin(top_countries.index)], x='country',␣
 ↪hue='type')
plt.title('Movies vs TV Shows by Country')
plt.xticks(rotation=90)
plt.show()
```

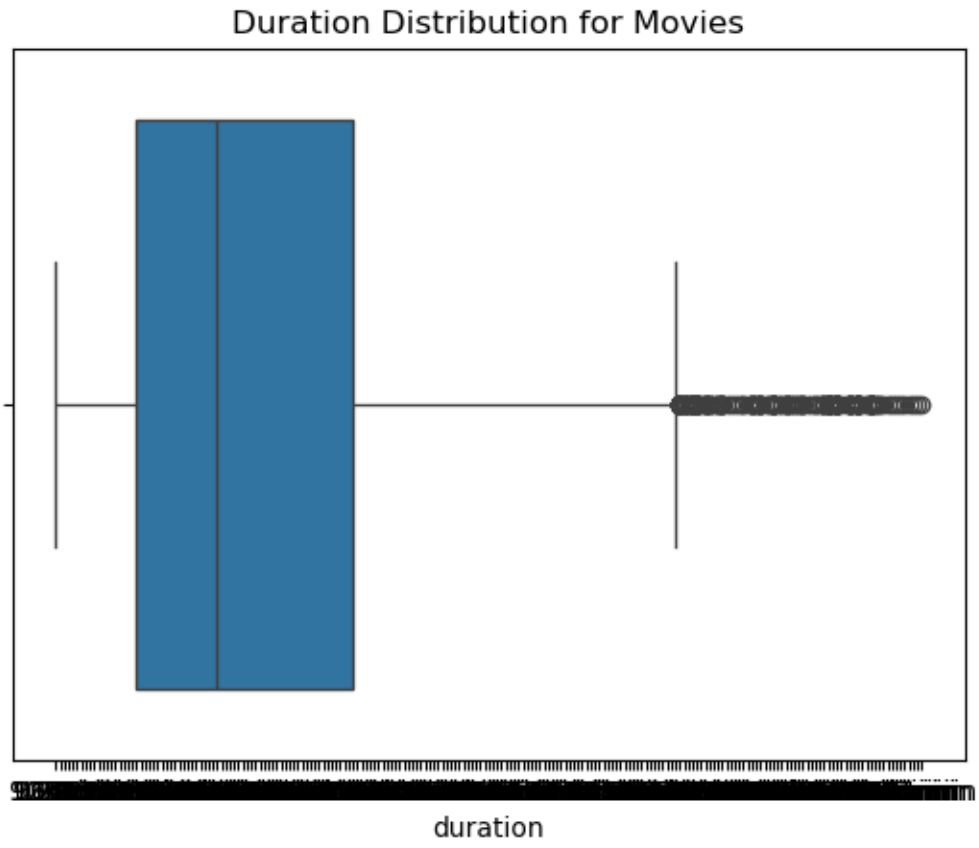


[30]:
```
df['director'].value_counts().head(10).plot(kind='bar', figsize=(10,5))
plt.title('Top 10 Directors on Netflix')
plt.show()
```



[32]:
```
df_clean = df.dropna(subset=['director', 'country'])
```

[48]:
```
sns.boxplot(data=df[df['type']=='Movie'], x='duration')
plt.title('Duration Distribution for Movies')
plt.show()
```



Business Insights:Based on the visualizations and analysis:

Movies vs TV Shows: If you find Netflix producing more TV shows in recent years, it indicates that serialized content is their focus. Regional Focus: Countries like the US, India, and the UK dominate Netflix's catalog. Netflix may need to increase content from underrepresented regions. Popular Genres: If thrillers, documentaries, or dramas are dominant, Netflix can focus more on producing these genres.

Recommendations Here are some recommendations based on your analysis:

Increase Regional Content: Focus on producing more local content in countries like Brazil, Japan, or Spain, which could lead to better audience retention and growth in these regions.

Boost TV Show Production: The shift in focus towards TV shows is working well. Netflix should continue investing in quality TV shows, especially multi-seasonal series.

Leverage Popular Genres: Popular genres like thrillers, dramas, and documentaries should be prioritized. Creating genre-specific marketing campaigns can boost viewership.

Director and Cast Partnerships: Collaborating with top directors and actors (frequent in Netflix's catalog) could help build brand loyalty and ensure consistent viewership.

[ ]: