



# Netflix Data Analysis Project

Exploring content trends using Python, Pandas, and Seaborn.

This project analyzes Netflix titles to uncover trends by content type, genres, country, and release patterns. It uses visualizations like bar charts, line plots, pie charts, and stacked bars to draw insights.

```
In [1]: import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [3]: df=pd.read_csv(r"C:\Users\Lenovo\Downloads\netflix_titles.csv.zip")
```

```
In [4]: df
```

Out[4]:

[illegible]

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozes Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

8807 rows × 12 columns

```
In [5]: df.isnull().sum()
```

```
Out[5]: 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
```

```
In [15]: df_clean=df.drop(columns='show_id')
```

```
In [16]: df=df_clean
```

```
In [17]: df
```

Out[17]:

	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
<b>0</b>	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries
<b>1</b>	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
<b>2</b>	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...
<b>3</b>	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
<b>4</b>	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 ...
...	...	...	...	...	...	...	...	...	...	...
<b>8802</b>	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers
<b>8803</b>	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies
<b>8804</b>	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies
<b>8805</b>	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies

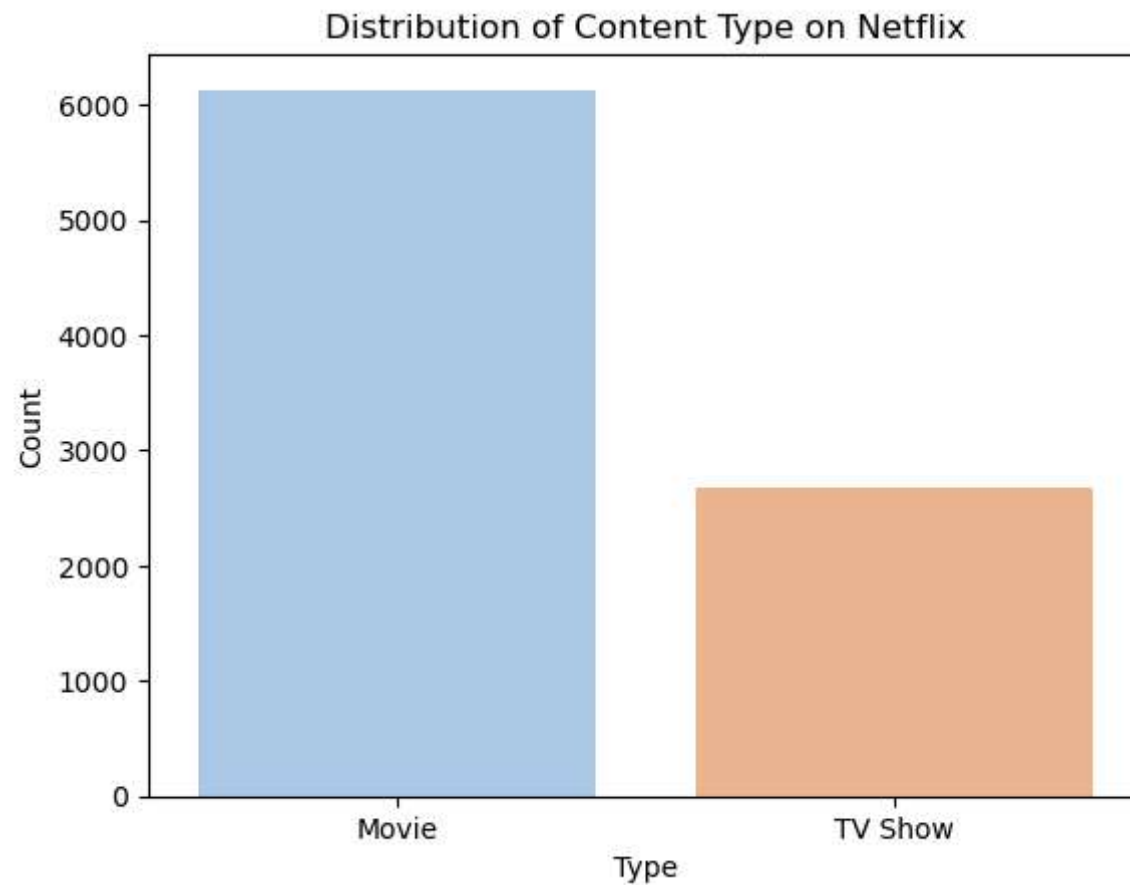
	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
8806	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals

8807 rows × 10 columns

## Insight 1: Content Type Distribution

We compare the number of Movies and TV Shows available on Netflix.

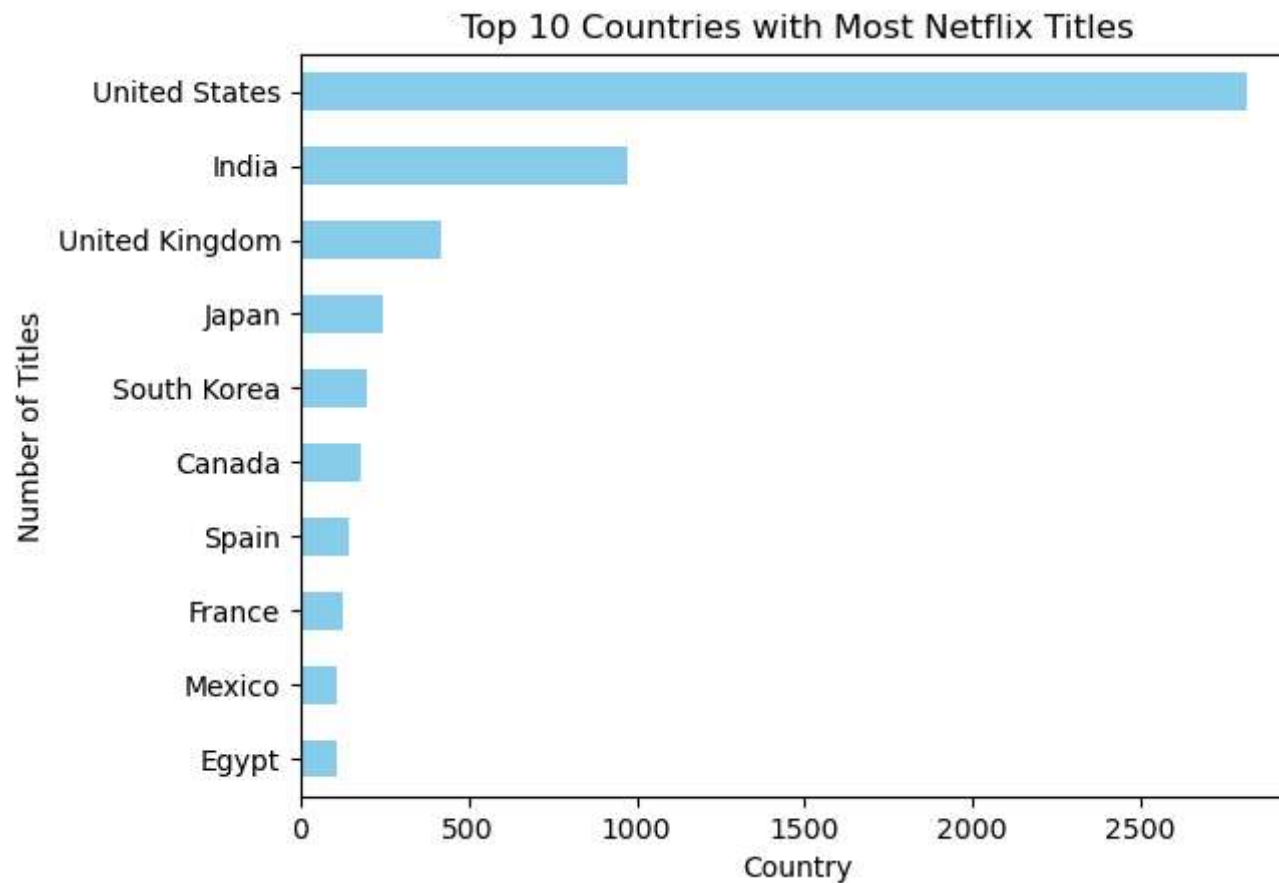
```
In [39]: sns.countplot(data=df, x='type', hue='type', palette='pastel', legend=False)
plt.title("Distribution of Content Type on Netflix")
plt.xlabel("Type")
plt.ylabel("Count")
plt.show()
```



## Insight 2: Top 10 Countries with Most Titles

This insight highlights which countries contribute the most to Netflix's content library, indicating regional production strength.

```
In [29]: top_countries=df['country'].value_counts().dropna().head(10)
top_countries.plot(kind='barh',color='skyblue')
plt.title("Top 10 Countries with Most Netflix Titles")
plt.ylabel("Number of Titles")
plt.xlabel("Country")
plt.gca().invert_yaxis()
plt.show()
```



## Insight 3: Top 5 Genres on Netflix

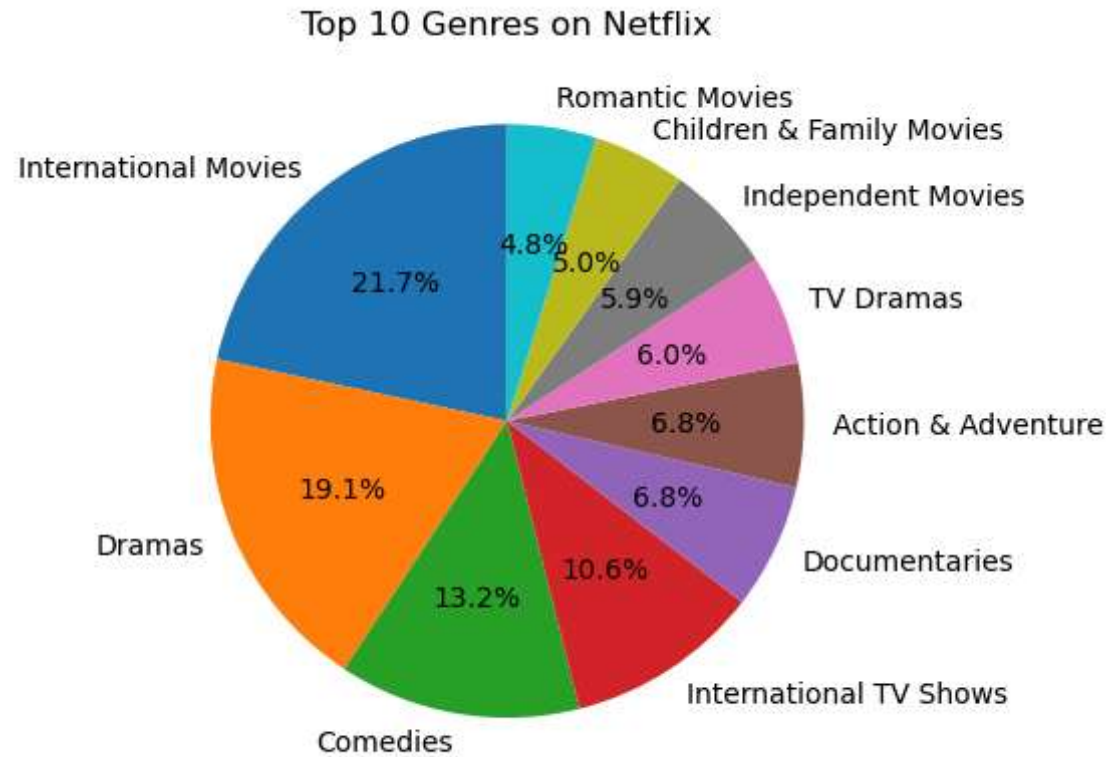
This chart helps us identify the most popular content genres on Netflix by frequency of appearance.

```
In [30]: genres = df['listed_in'].dropna().str.split(', ')
flat_genres = [genre for sublist in genres for genre in sublist]
genre_series = pd.Series(flat_genres).value_counts().head(10)

genre_series.plot(kind='pie', autopct='%1.1f%%', startangle=90, color=['skyblue', 'gold', 'orchid', 'lightcoral', 'lightgreen'])
plt.title("Top 10 Genres on Netflix")
```



```
plt.ylabel("")  
plt.show()
```

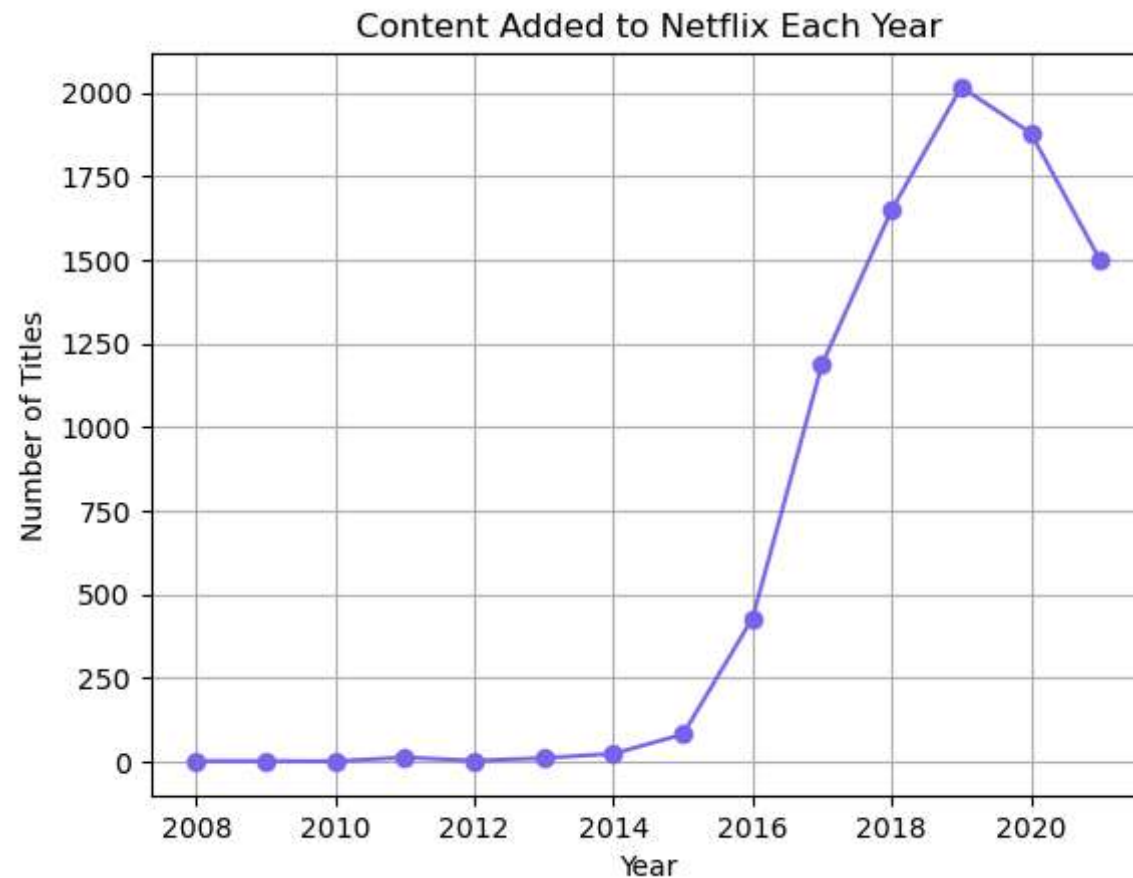


## Insight 4: Content Added Over the Years

We examine how Netflix's content library has grown year by year based on the date content was added to the platform.

```
In [34]: df['date_added'] = df['date_added'].str.strip()  
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')  
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', color='mediumslateblue')  
plt.title("Content Added to Netflix Each Year")
```

```
plt.xlabel("Year")
plt.ylabel("Number of Titles")
plt.grid(True)
plt.show()
```

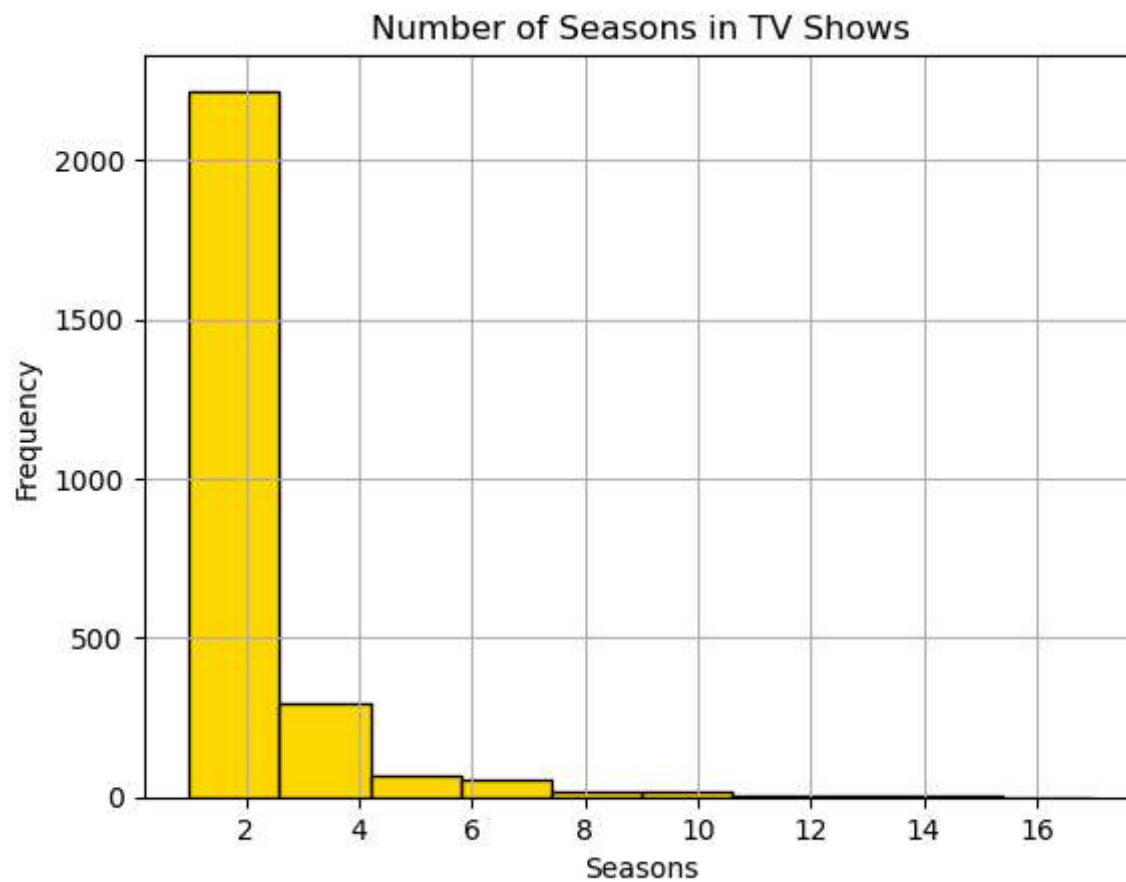


## Insight 5: Seasons Distribution in TV Shows

This histogram shows how many TV Shows have 1, 2, or more seasons, giving insight into Netflix's focus on limited series vs long-running shows.

```
In [37]: tv_shows = df[df['type'] == 'TV Show'].copy()
tv_shows['seasons'] = tv_shows['duration'].str.extract('(\d+)').astype(float)

plt.hist(tv_shows['seasons'].dropna(), bins=10, color='gold', edgecolor='black')
plt.title("Number of Seasons in TV Shows")
plt.xlabel("Seasons")
plt.ylabel("Frequency")
plt.grid(True)
plt.show()
```



## Insight 5: Seasons Distribution in TV Shows

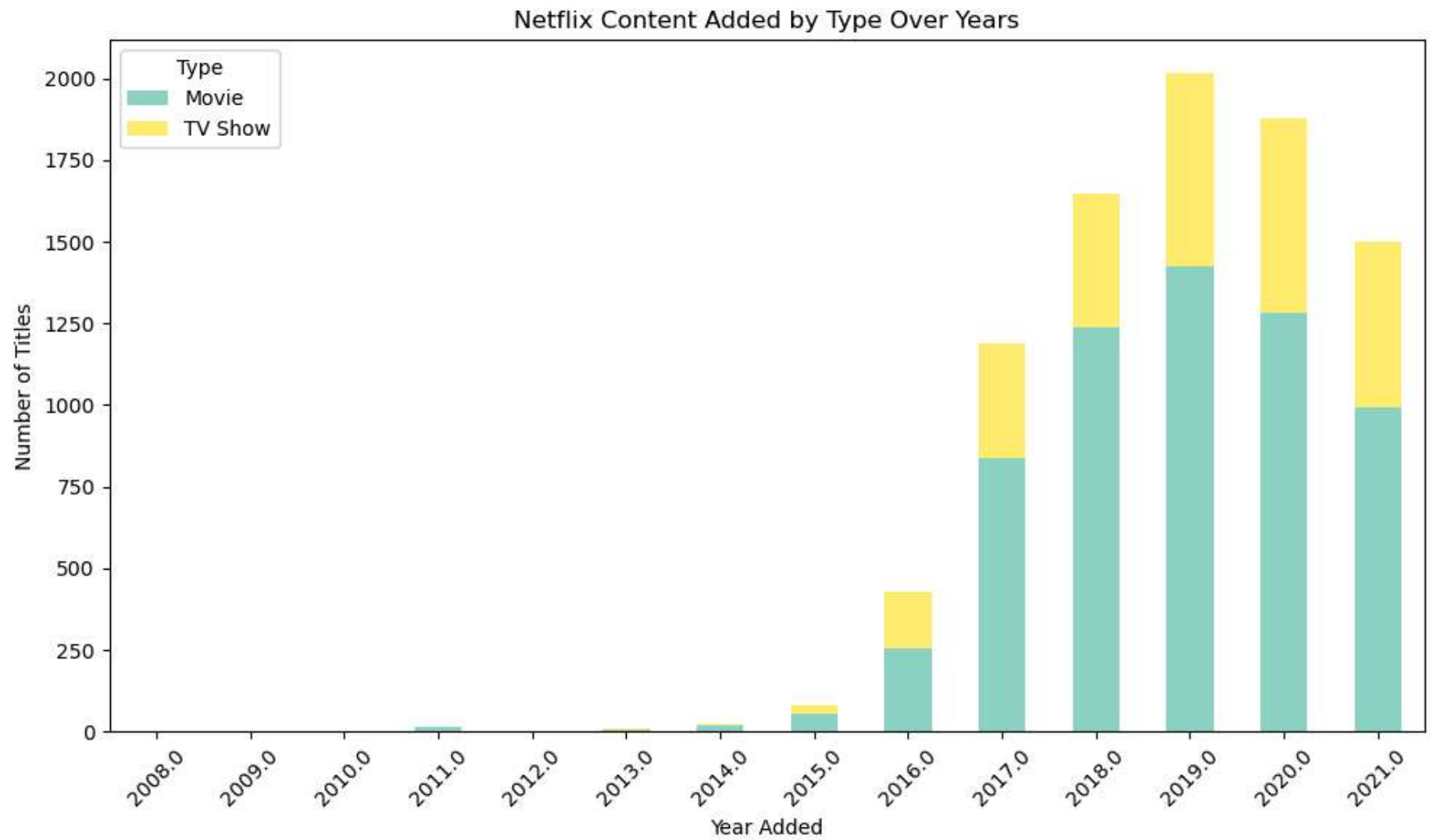
This histogram shows how many TV Shows have 1, 2, or more seasons, giving insight into Netflix's focus on limited series vs long-running shows.

```
In [38]: # Convert 'date_added' to datetime and extract year
df['date_added'] = pd.to_datetime(df['date_added'])
df['year_added'] = df['date_added'].dt.year

# Group by year and type, then count
year_type_counts = df.groupby(['year_added', 'type']).size().unstack(fill_value=0)

# Plot as stacked bar
year_type_counts.plot(kind='bar', stacked=True, colormap='Set3', figsize=(10, 6))

plt.title("Netflix Content Added by Type Over Years")
plt.xlabel("Year Added")
plt.ylabel("Number of Titles")
plt.xticks(rotation=45)
plt.legend(title='Type')
plt.tight_layout()
plt.show()
```



## Final Insights Summary

- Netflix has more Movies than TV Shows, favoring short-form content.
- The US dominates content contribution, followed by India and UK.
- Dramas and Comedies are the most popular genres.

- Content addition peaked in 2019–2020, showing Netflix's rapid growth.
- Most TV Shows are single-season, highlighting a trend toward mini-series.
- Stacked chart shows consistent growth of both Movies and TV Shows over time.

---

Project completed using: Pandas, Matplotlib, Seaborn.