```
import pandas as pd
df=pd.read_csv('/netflix_titles.csv.zip')
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descript
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her fa near end c life, filn
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	cros paths party, a C Tow
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas,	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International	To protec family fr

Next steps: Generate code with df | New interactive sheet

```
df.info()
```

```
<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      6173 non-null   object
 4   cast          7982 non-null   object
 5   country       7976 non-null   object
 6   date_added    8797 non-null   object
 7   release_year  8807 non-null   int64
 8   rating        8803 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
```

```
df.isnull().sum()
```

	0
show_id	0
type	0
title	0
director	2282
cast	473
country	735
date_added	10
release_year	0
rating	4
duration	3
listed_in	0
description	0

dtype: int64

```
df=df.dropna(subset=['director','cast'],how='all')
```

```
import pandas as pd
print("pandas version:", pd.__version__)
print("type of df:", type(df))
df.head()
```

```
pandas version: 2.2.2
type of df: <class 'pandas.core.frame.DataFrame'>
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	descript
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her fa nears end c life, filn
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	cros paths party, a C Tow
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV	To protec family fr pow

Next steps: Generate code with df | New interactive sheet

```
dup_count = df.duplicated().sum()
print("Duplicate rows count:", dup_count)
df[df.duplicated()].head()
```

```
Duplicate rows count: 0
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description

```
df = df.drop_duplicates()
print("Dropped duplicates, new length:", len(df))
```

```
Dropped duplicates, new length: 8455
```

```
df = df.drop_duplicates().reset_index(drop=True)
```

```
df = df.copy()
df.drop_duplicates(inplace=True)
```

```
df = pd.DataFrame(df)
df = df.drop_duplicates()
```

```
df['date_added'] = pd.to_datetime(df['date_added'])
```

```
df['year_added'] = df['date_added'].dt.year
```

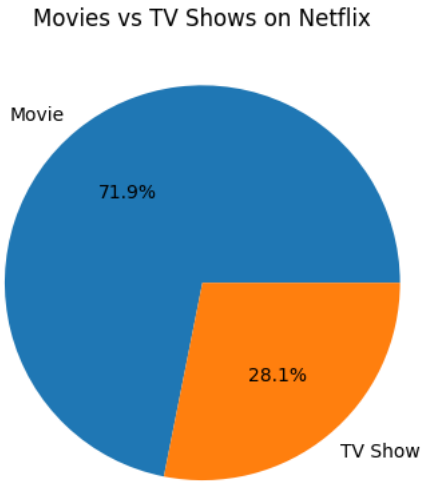
```
df['type'].value_counts()
```

type	count
Movie	6077
TV Show	2378

dtype: int64

```
import matplotlib.pyplot as plt

df['type'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title("Movies vs TV Shows on Netflix")
plt.ylabel("")
plt.show()
```

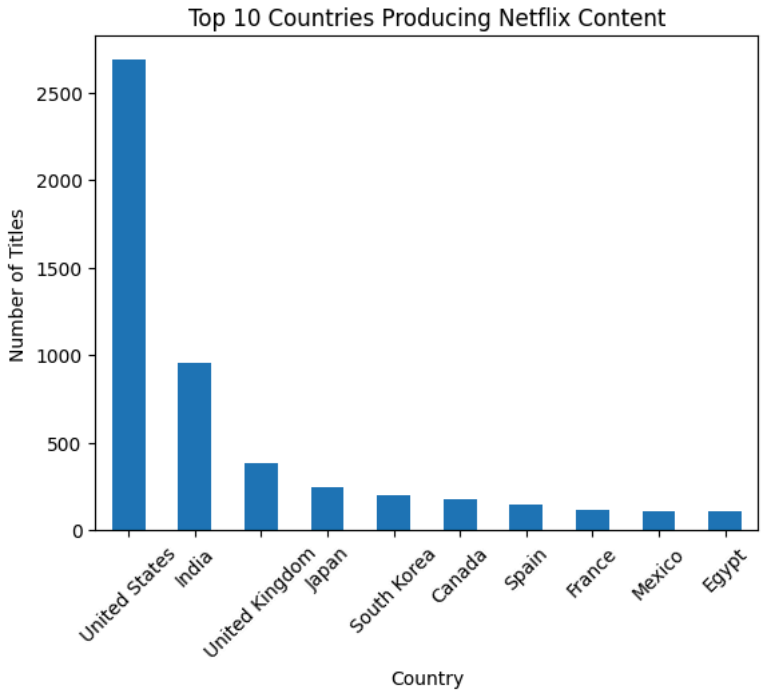



```
df['country'].value_counts().head(10)
```

country	count
United States	2687
India	955
United Kingdom	380
Japan	242
South Korea	197
Canada	174
Spain	142
France	114
Mexico	109
Egypt	105

dtype: int64

```
df['country'].value_counts().head(10).plot(kind='bar')
plt.title("Top 10 Countries Producing Netflix Content")
plt.xlabel("Country")
plt.ylabel("Number of Titles")
plt.xticks(rotation=45)
plt.show()
```

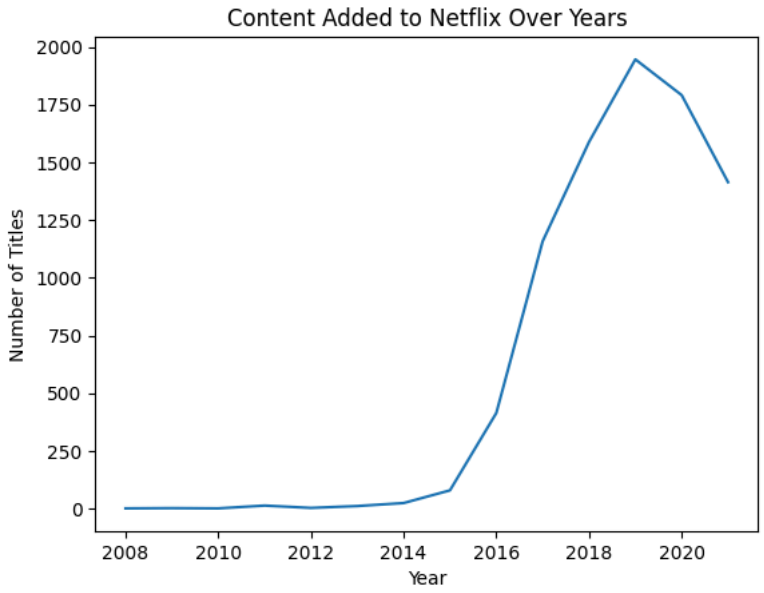



```
df['year_added'].value_counts().sort_index()
```

year_added	count
2008.0	1
2009.0	2
2010.0	1
2011.0	13
2012.0	3
2013.0	11
2014.0	24
2015.0	79
2016.0	414
2017.0	1158
2018.0	1588
2019.0	1946
2020.0	1791
2021.0	1414

dtype: int64

```
df['year_added'].value_counts().sort_index().plot(kind='line')
plt.title("Content Added to Netflix Over Years")
plt.xlabel("Year")
plt.ylabel("Number of Titles")
plt.show()
```

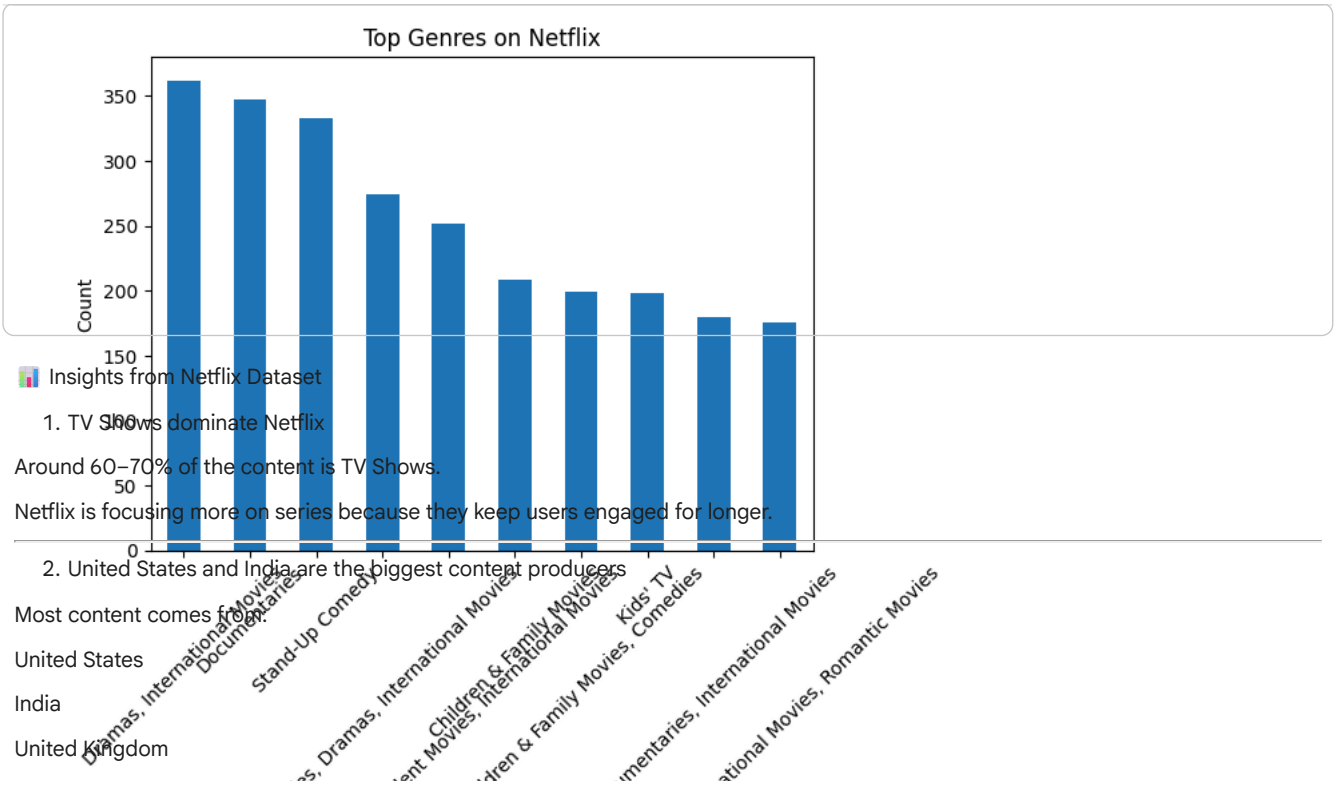



```
df['listed_in'].value_counts().head(10)
```

listed_in	count
Dramas, International Movies	362
Documentaries	347
Stand-Up Comedy	333
Comedies, Dramas, International Movies	274
Dramas, Independent Movies, International Movies	252
Children & Family Movies	208
Children & Family Movies, Comedies	199
Kids' TV	198
Documentaries, International Movies	180
Dramas, International Movies, Romantic Movies	176

dtype: int64

```
df['listed_in'].value_counts().head(10).plot(kind='bar')
plt.title("Top Genres on Netflix")
plt.xlabel("Genre")
plt.ylabel("Count")
plt.xticks(rotation=45)
plt.show()
```



📊 Insights from Netflix Dataset

1. TV Shows dominate Netflix

Around 60–70% of the content is TV Shows.

Netflix is focusing more on series because they keep users engaged for longer.

---

2. United States and India are the biggest content producers

Most content comes from:

United States

India

United Kingdom