

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

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

	show_id	type	title	director	\
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	
1	s2	TV Show	Blood & Water	NaN	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	NaN	
4	s5	TV Show	Kota Factory	NaN	

	cast	country	\
0	NaN	United States	
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	
3	NaN	NaN	
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	

	date_added	release_year	rating	duration	\
0	September 25, 2021	2020	PG-13	90 min	
1	September 24, 2021	2021	TV-MA	2 Seasons	
2	September 24, 2021	2021	TV-MA	1 Season	
3	September 24, 2021	2021	TV-MA	1 Season	
4	September 24, 2021	2021	TV-MA	2 Seasons	

	listed_in	\
0	Documentaries	
1	International TV Shows, TV Dramas, TV Mysteries	
2	Crime TV Shows, International TV Shows, TV Act...	
3	Docuseries, Reality TV	
4	International TV Shows, Romantic TV Shows, TV ...	

	description
0	As her father nears the end of his life, filmm...
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...

```
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.describe(include='object')
```

```

count      show_id  type  title      director      cast  \
unique      8807      2    8807      4528      7692
top      s8807  Movie  Zubaan  Rajiv Chilaka  David Attenborough
freq          1    6131          1          19          19

```

```

count      country      date_added  rating  duration  \
unique          748          1767      17          220
top      United States  January 1, 2020  TV-MA  1 Season
freq          2818          109    3207    1793

```

```

count      listed_in  \
unique          514
top      Dramas, International Movies
freq          362

```

```

count      description
unique          8775
top      Paranormal activity at a lush, abandoned prope...
freq          4

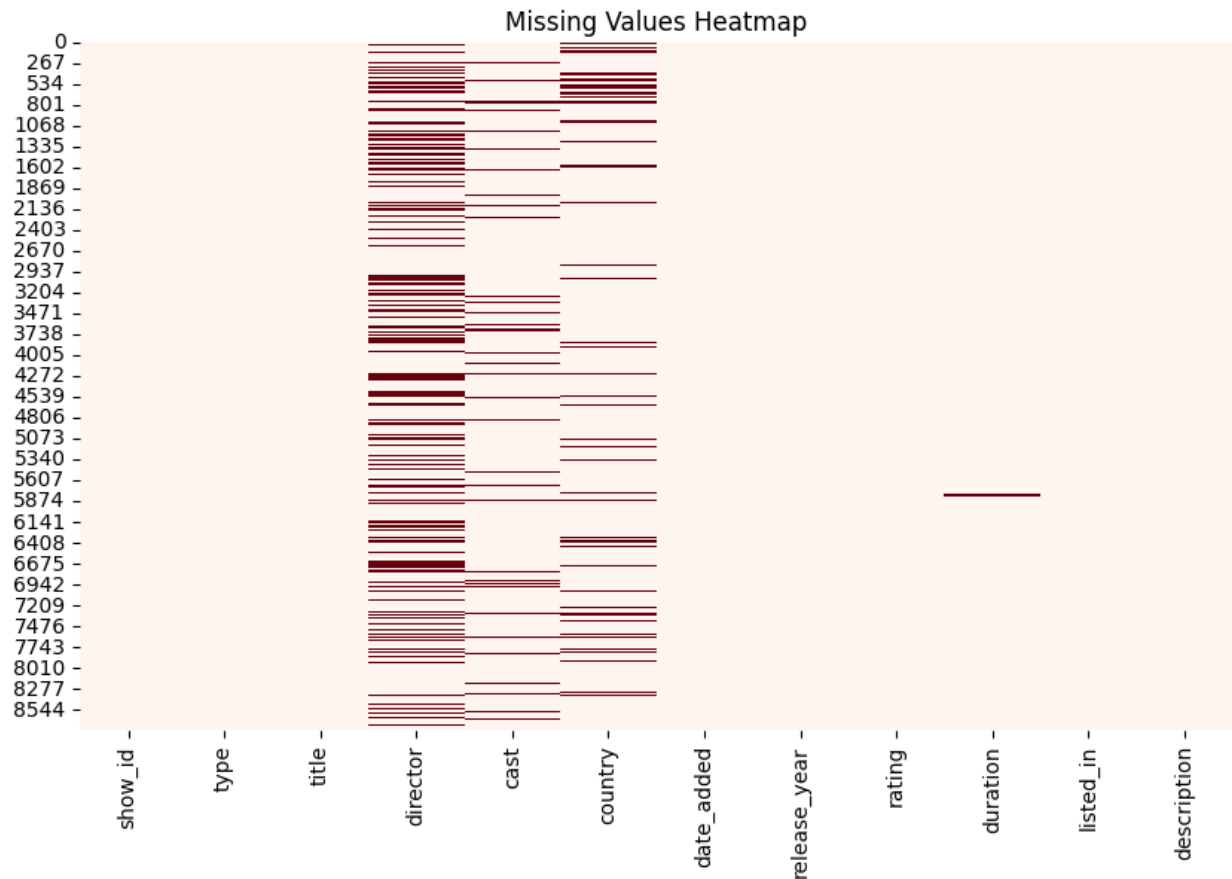
```

```
# Checking for missing values with the help of heatmap
```

```

plt.figure(figsize=(10,6))
sns.heatmap(df.isnull(), cbar=False, cmap="Reds")
plt.title("Missing Values Heatmap")
plt.show()

```



```
# Replacing blanks/missing values with Unknown/not specified/ not
rated etc...
df['director'] = df['director'].fillna('Unknown')
df['cast'] = df['cast'].fillna('Not Specified')
df['country'] = df['country'].fillna('Not Specified')
df['rating'] = df['rating'].fillna('Not Rated')

# Strip extra spaces in date strings
df['date_added'] = df['date_added'].astype(str).strip()

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

# Extracting year and month from date_added
df['year_added'] = df['date_added'].dt.year
df['month_added'] = df['date_added'].dt.month

df.head()
```

	show_id	type	title	director	
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	\
1	s2	TV Show	Blood & Water	Unknown	
2	s3	TV Show	Ganglands	Julien Leclercq	
3	s4	TV Show	Jailbirds New Orleans	Unknown	

4	s5	TV Show	Kota Factory	Unknown
			cast	country \
0			Not Specified	United States
1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...			South Africa
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...			Not Specified
3			Not Specified	Not Specified
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...			India
	date_added	release_year	rating	duration \
0	2021-09-25	2020	PG-13	90 min
1	2021-09-24	2021	TV-MA	2 Seasons
2	2021-09-24	2021	TV-MA	1 Season
3	2021-09-24	2021	TV-MA	1 Season
4	2021-09-24	2021	TV-MA	2 Seasons
			listed_in \	
0			Documentaries	
1	International TV Shows, TV Dramas, TV Mysteries			
2	Crime TV Shows, International TV Shows, TV Act...			
3	Docuseries, Reality TV			
4	International TV Shows, Romantic TV Shows, TV ...			
			description	year_added
month_added				
0	As her father nears the end of his life, filmm...			2021.0
9.0				
1	After crossing paths at a party, a Cape Town t...			2021.0
9.0				
2	To protect his family from a powerful drug lor...			2021.0
9.0				
3	Feuds, flirtations and toilet talk go down amo...			2021.0
9.0				
4	In a city of coaching centers known to train I...			2021.0
9.0				
# converting year_added and month_added float to int				
df['year_added'] = df['year_added'].astype('Int64')				
df['month_added'] = df['month_added'].astype('Int64')				
df.head()				
	show_id	type	title	director \
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson
1	s2	TV Show	Blood & Water	Unknown
2	s3	TV Show	Ganglands	Julien Leclercq
3	s4	TV Show	Jailbirds New Orleans	Unknown
4	s5	TV Show	Kota Factory	Unknown
			cast	country \
0			Not Specified	United States

1	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa
2	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	Not Specified
3	Not Specified	Not Specified
4	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India

	date_added	release_year	rating	duration	\
0	2021-09-25	2020	PG-13	90 min	
1	2021-09-24	2021	TV-MA	2 Seasons	
2	2021-09-24	2021	TV-MA	1 Season	
3	2021-09-24	2021	TV-MA	1 Season	
4	2021-09-24	2021	TV-MA	2 Seasons	

	listed_in	\
0	Documentaries	
1	International TV Shows, TV Dramas, TV Mysteries	
2	Crime TV Shows, International TV Shows, TV Act...	
3	Docuseries, Reality TV	
4	International TV Shows, Romantic TV Shows, TV ...	

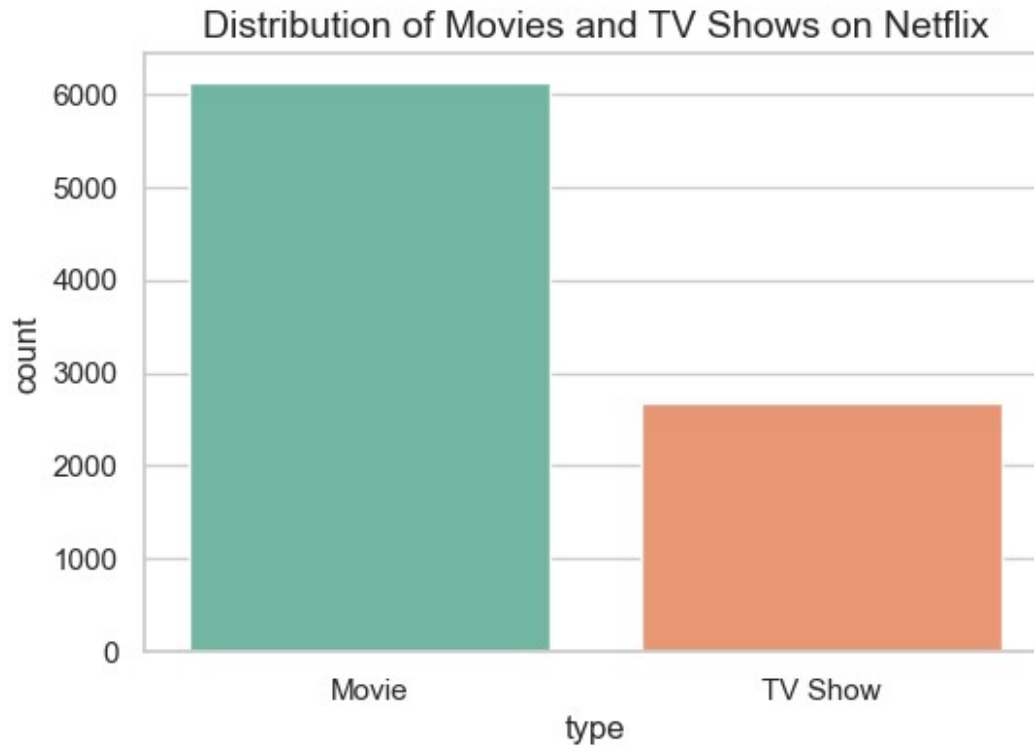
	description	year_added
month_added		
0	As her father nears the end of his life, filmm...	2021
9		
1	After crossing paths at a party, a Cape Town t...	2021
9		
2	To protect his family from a powerful drug lor...	2021
9		
3	Feuds, flirtations and toilet talk go down amo...	2021
9		
4	In a city of coaching centers known to train I...	2021
9		

```
# Showing the Distribution of Movies and TV Shows on Netflix
sns.set(style="whitegrid")
```

```
plt.figure(figsize=(6,4))
sns.countplot(data=df, x='type' , hue='type' , palette='Set2')
```

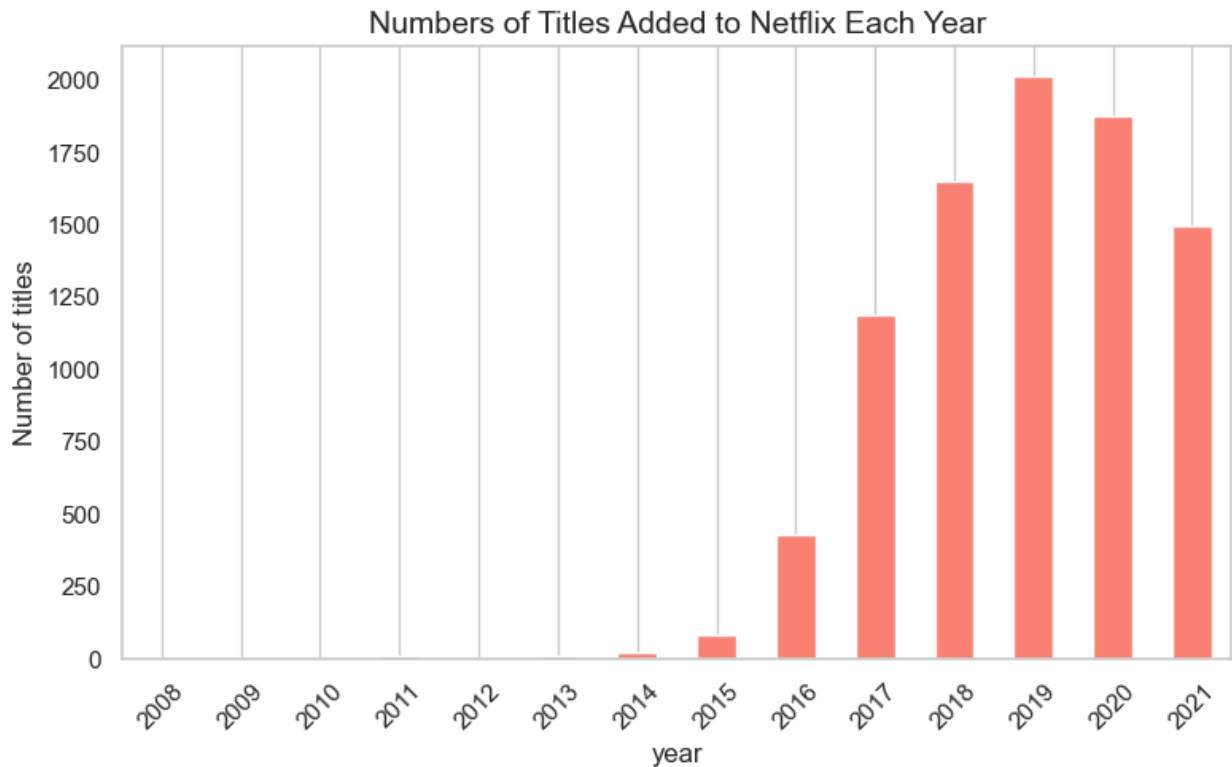
```
plt.title("Distribution of Movies and TV Shows on Netflix",
fontsize=14)
plt.xlabel("type", fontsize=12)
plt.ylabel("count", fontsize=12)
```

```
plt.show()
```



```
# Showing the Numbers of Titles Added to Netflix Each Year
plt.figure(figsize=(9,5))
df['year_added'].value_counts().sort_index().plot(kind='bar' ,
color='salmon')

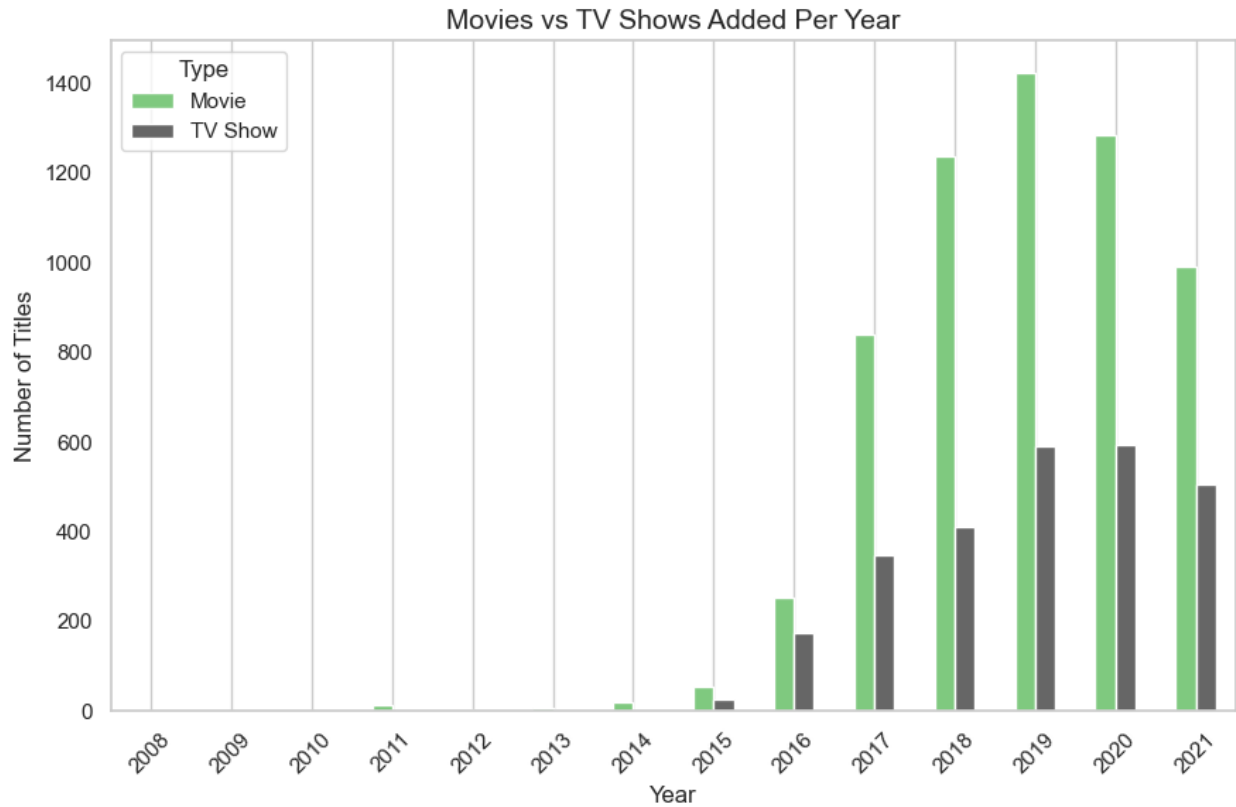
plt.title("Numbers of Titles Added to Netflix Each Year", fontsize=14)
plt.xlabel("year", fontsize=12)
plt.ylabel("Number of titles", fontsize=12)
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.show()
```



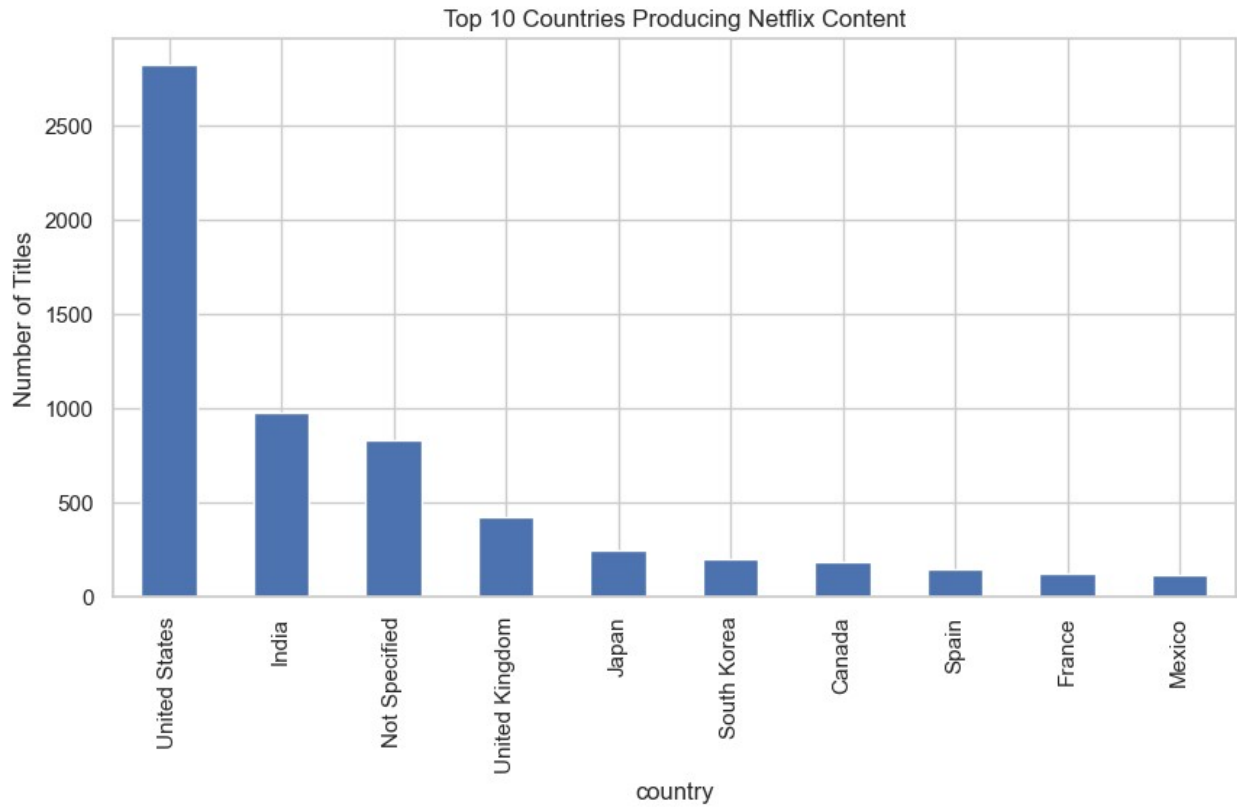
```
# Showing the Movies vs TV Shows Added Per Year
year_type = df.groupby(['year_added', 'type']).size().unstack()

year_type.plot(kind='bar', stacked=False, figsize=(9,6),
               colormap='Accent')

plt.title("Movies vs TV Shows Added Per Year", fontsize=14)
plt.xlabel("Year", fontsize=12)
plt.ylabel("Number of Titles", fontsize=12)
plt.xticks(rotation=45)
plt.legend(title="Type")
plt.grid(axis='y')
plt.tight_layout()
plt.show()
```



```
#Showing the Top 10 Countries Producing Netflix Content
top_countries= df['country'].value_counts().head(10)
top_countries.plot(kind='bar', title= 'Top 10 Countries Producing
Netflix Content', figsize=(10,5))
plt.xlabel("country")
plt.ylabel("Number of Titles")
plt.show()
```

```
# Showing the Top 10 Director on Netflix
top_directors = df['director'].value_counts().head(10)
top_directors.plot(kind='barh', title= 'Top 10 Director on Netflix')
plt.xlabel("Number of Titles")
plt.show()
```

