

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

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

	show_id	type	title	director	cast	country	date_added	release
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	NaN	September 24, 2021	

Next steps:

[Generate code with data](#)[New interactive sheet](#)

```
data.info()
data.isnull().sum()
```

```

<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

```

	0
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

```

for column in data.select_dtypes(include=np.number).columns:
    data[column].fillna(data[column].mean(), inplace=True)

for column in data.select_dtypes(include='object').columns:
    data[column].fillna(data[column].mode()[0], inplace=True)

data.isnull().sum()

```

```
/tmp/ipykernel_267/653201675.py:2: FutureWarning: A value is trying to be set
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.

```
data[column].fillna(data[column].mean(), inplace=True)
/tmp/ipykernel_267/653201675.py:5: FutureWarning: A value is trying to be set
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.

```
data[column].fillna(data[column].mode()[0], inplace=True)
```

	0
show_id	0
type	0
title	0
director	0
cast	0
country	0
date_added	0
release_year	0
rating	0
duration	0
listed_in	0
description	0

dtype: int64

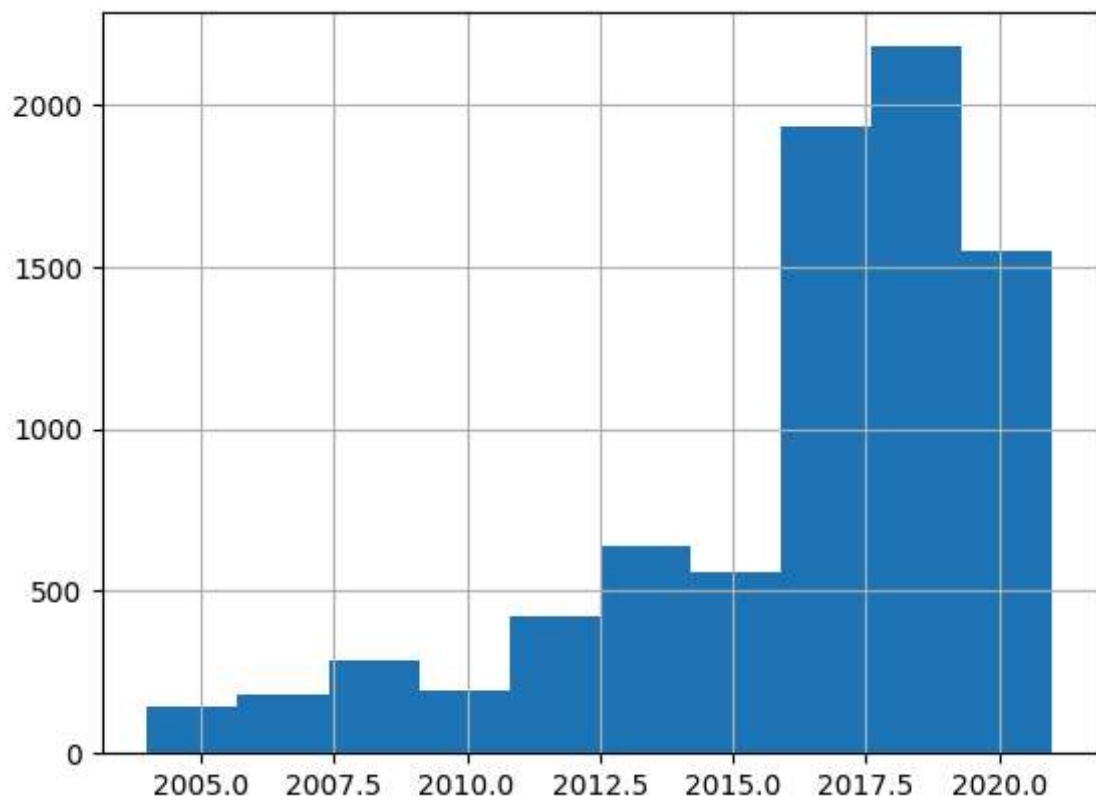
```
def remove_outliers(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    df = df[(df[column] >= lower) & (df[column] <= upper)]
    return df
```

```
for column in data.select_dtypes(include=np.number).columns:
    data = remove_outliers(data, column)
```

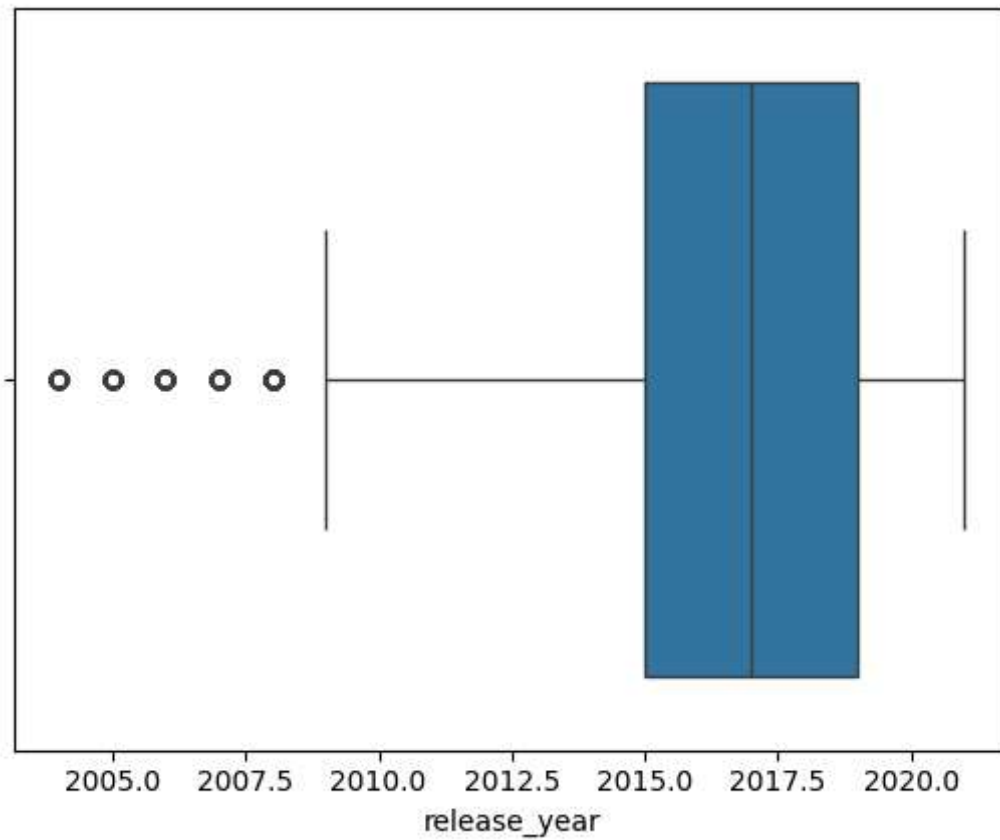
```
data.shape
```

(8088, 12)

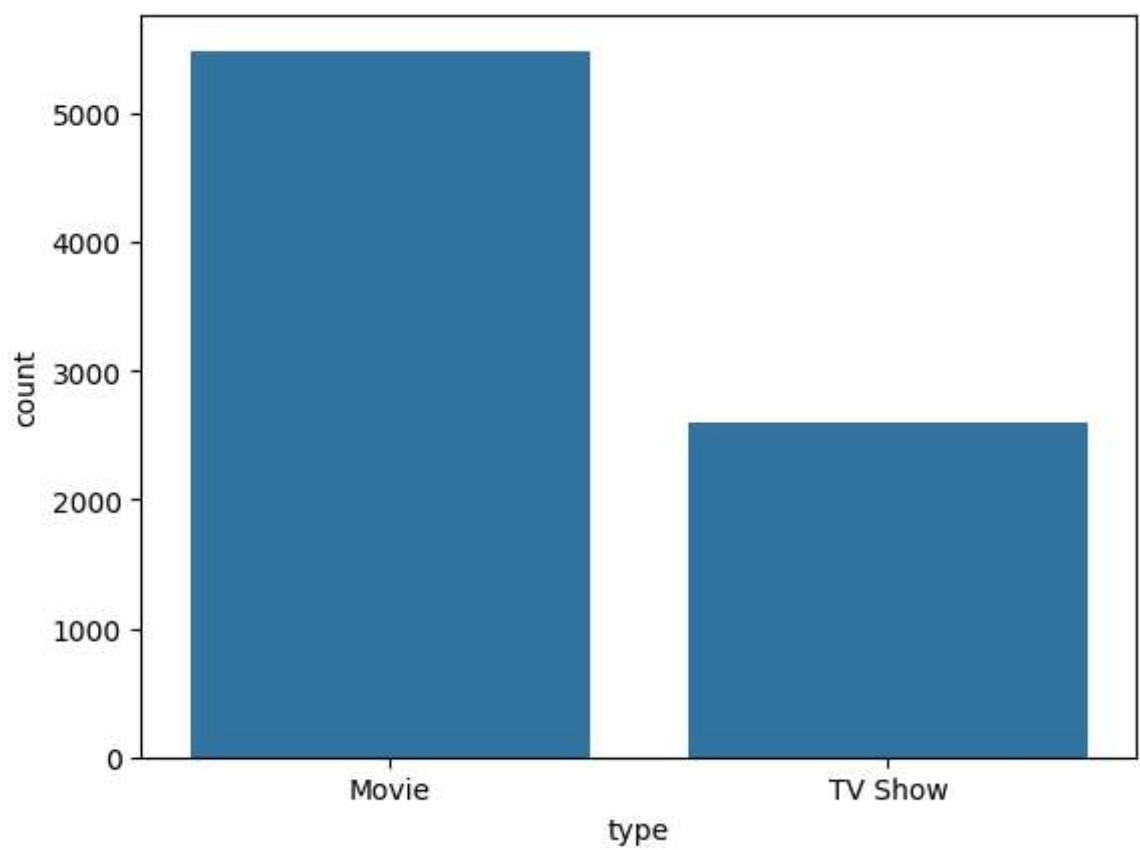
```
plt.figure()  
data['release_year'].hist()  
plt.show()
```



```
plt.figure()  
sns.boxplot(x=data['release_year'])  
plt.show()
```



```
plt.figure()  
sns.countplot(x=data['type'])  
plt.show()
```



```
plt.figure()  
data.corr(numeric_only=True)  
sns.heatmap(data.corr(numeric_only=True), annot=True)
```