4zcj6qt6f

March 10, 2025

1 Business Case Study - Netflix

2 Business Problem

Analyze the data and generate insights that could help Netflix ijn deciding which type of shows/movies to produce and how they can grow the business in different countries

3 About NETFLIX

Netflix is one of the most popular media and video streaming platforms. They have over 10000 movies or tv shows available on their platform, as of mid-2021, they have over 222M Subscribers globally. This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc.

4 Objectives of the Project

- Perform EDA on the given dataset and find insights.
- Provide Useful Insights and Business recommendations that can help the business to grow.

5 1) Importing Libraries and the Netflix dataset

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

[ ]:
```
!gdown 1r1xysMLi3PED70IIAaFI0CLZOT1JxjB4
```

```
Downloading…
From: https://drive.google.com/uc?id=1r1xysMLi3PED70IIAaFI0CLZOT1JxjB4
To: /content/netflix.csv
100% 3.40M/3.40M [00:00<00:00, 189MB/s]
```

[ ]:
```
netflix = pd.read_csv('netflix.csv')
netflix.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…
```

Lets check the shape and data distribution of the data set

[ ]:
```
netflix.shape
```

[ ]:
```
(8807, 12)
```

There are 8807 rows and 12 columns in the given dataset Lets check the info

[ ]:
```
netflix.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
```

[ ]:
```
netflix.nunique()
```

[ ]:
```
show_id         8807
type               2
title           8807
director        4528
cast            7692
country          748
date_added      1767
release_year      74
rating            17
duration         220
listed_in        514
description     8775
dtype: int64
```

[ ]:
```
netflix.describe()
```

[ ]:
```
       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
```

[ ]:
```
netflix.describe(include=object)
```

[ ]:
```
       show_id   type                  title       director  \
count     8807   8807                   8807           6173
unique    8807      2                   8807           4528
top         s1  Movie  Dick Johnson Is Dead  Rajiv Chilaka
freq         1   6131                      1             19

                       cast        country       date_added rating  duration  \
count                  7982           7976             8797   8803      8804
unique                 7692            748             1767     17       220
top     David Attenborough  United States  January 1, 2020  TV-MA  1 Season
freq                     19           2818              109   3207      1793

                           listed_in  \
count                           8807
unique                           514
top     Dramas, International Movies
freq                             362

                                                 description
count                                                   8807
unique                                                  8775
top     Paranormal activity at a lush, abandoned prope…
freq                                                       4
```

6 2) Statistical Summary

Most common country : United States (2,818 shows)

Most frequent rating : TV-MA (3,207 shows)

Most common category : Dramas, International Movies

Earliest release year : 1925

Latest release year : 2021

7 3) Non graphical analysis

[ ]:
```
netflix['type'].value_counts()
```

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

There are only two types of contents on Netflex

- TV Shows
- Movies

Moviews are almost 3 times higher than TV shows

[ ]:
```
netflix['director'].value_counts()
```

[ ]:
```
director
Rajiv Chilaka                     19
Raúl Campos, Jan Suter            18
Marcus Raboy                      16
Suhas Kadav                       16
Jay Karas                         14
                                  ..
Raymie Muzquiz, Stu Livingston     1
Joe Menendez                       1
Eric Bross                         1
Will Eisenberg                     1
Mozez Singh                        1
Name: count, Length: 4528, dtype: int64
```

Rajiv Chillaka has directed most number of Movies/TV Shows

[ ]:
```
netflix['country'].value_counts()
```

[ ]:
```
country
United States                                2818
India                                         972
United Kingdom                                419
Japan                                         245
South Korea                                   199
                                             …
Romania, Bulgaria, Hungary                      1
Uruguay, Guatemala                              1
France, Senegal, Belgium                        1
Mexico, United States, Spain, Colombia          1
United Arab Emirates, Jordan                    1
Name: count, Length: 748, dtype: int64
```

The United States has the highest count of content (2818)

[ ]:
```
netflix['date_added'].value_counts()
```

[ ]:
```
date_added
January 1, 2020      109
November 1, 2019      89
March 1, 2018         75
December 31, 2019     74
October 1, 2018       71
                     …
December 4, 2016       1
```

```
November 21, 2016      1
November 19, 2016      1
November 17, 2016      1
January 11, 2020       1
Name: count, Length: 1767, dtype: int64
```

Most Content were added on January 1, 2020.

[ ]:
```
(netflix['release_year'].min(),netflix['release_year'].max())
```

[ ]:
```
(1925, 2021)
```

The oldest content on Netflix was released in 1925. The newest content was released in 2021.

[ ]:
```
netflix['rating'].value_counts()
```

[ ]:
```
rating
TV-MA       3207
TV-14       2160
TV-PG        863
R            799
PG-13        490
TV-Y7        334
TV-Y         307
PG           287
TV-G         220
NR            80
G             41
TV-Y7-FV       6
NC-17          3
UR             3
74 min         1
84 min         1
66 min         1
Name: count, dtype: int64
```

TV-MA (3207) is the most common rating.

[ ]:
```
netflix.groupby('type')['duration'].value_counts()
```

[ ]:
```
type     duration
Movie    90 min        152
         93 min        146
         94 min        146
         97 min        146
         91 min        144
                      …
TV Show  13 Seasons      3
```

```
          11 Seasons      2
          12 Seasons      2
          15 Seasons      2
          17 Seasons      1
Name: count, Length: 220, dtype: int64
```

We can see that duration of movies is in minutes and TV Shows in Seasons

[ ]:
```
netflix['listed_in'].value_counts()
```

[ ]:
```
listed_in
Dramas, International Movies                            362
Documentaries                                           359
Stand-Up Comedy                                         334
Comedies, Dramas, International Movies                  274
Dramas, Independent Movies, International Movies        252
                                                       ...
Kids' TV, TV Action & Adventure, TV Dramas                1
TV Comedies, TV Dramas, TV Horror                         1
Children & Family Movies, Comedies, LGBTQ Movies          1
Kids' TV, Spanish-Language TV Shows, Teen TV Shows        1
Cult Movies, Dramas, Thrillers                            1
Name: count, Length: 514, dtype: int64
```

Netflix has most contents on the genre - Dramas, International Movies

---

Observations

- There are two types of contents -
  1. Movies
  2. TV Shows
- Rajiv Chillaka has directed most number of Movies/TV Shows
- The United States has the highest count of content and India at Second Position
- Most content were added in the Year 2020
- The oldest content on Netflix was released in 1925, indicating that Netflix includes some very old or classic titles in its library. This suggests that Netflix offers a mix of historical and vintage content, likely appealing to audiences interested in classic films or shows. The newest content was released in 2021, showing that Netflix regularly updates its library with recent and contemporary titles.
- TV-MA (3207) is the most common rating, indicating that a significant portion of Netflix's content is intended for mature audiences only. This suggests that Netflix has a large selection of content that may contain adult themes, strong language, or graphic violence.
- Movies length is measured in minutes and TV Shows by Seasons
- Netflix has most contents on the genre - Dramas, International Movies

8 4) Data Cleaning

Lets check the null values in each

[ ]:
```
netflix.isna().sum()
```

[ ]:
```
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
```

9 4.1) Cleaning date added column

[ ]:
```
netflix['date_added'].isna().sum()
```

[ ]:
```
10
```

[ ]:
```
netflix.drop(netflix[netflix['date_added'].isna()].index, inplace=True)
```

For 'date_added' column, all values confirm to date format, So we can convert its data type from object to datetime

[ ]:
```
netflix['date_added'] = pd.to_datetime(netflix['date_added'].str.strip())
netflix['date_added']
```

[ ]:
```
0      2021-09-25
1      2021-09-24
2      2021-09-24
3      2021-09-24
4      2021-09-24
          ...
8802   2019-11-20
8803   2019-07-01
8804   2019-11-01
8805   2020-01-11
8806   2019-03-02
Name: date_added, Length: 8797, dtype: datetime64[ns]
```

2.2) Cleaning Rating Column

10 4.2) Cleaning Rating Column

[ ]:
```
netflix['rating'].isna().sum()
```

[ ]: 4

[ ]:
```
netflix_NaN_index=netflix[netflix['rating'].isna()].index
```

[ ]:
```
netflix.loc[netflix_NaN_index]
```

[ ]:
```
     show_id     type                                                 title  \
5989   s5990    Movie  13TH: A Conversation with Oprah Winfrey & Ava …
6827   s6828  TV Show                  Gargantia on the Verdurous Planet
7312   s7313  TV Show                                        Little Lunch
7537   s7538    Movie                              My Honor Was Loyalty

             director                                                 cast  \
5989              NaN                          Oprah Winfrey, Ava DuVernay
6827              NaN  Kaito Ishikawa, Hisako Kanemoto, Ai Kayano, Ka…
7312              NaN  Flynn Curry, Olivia Deeble, Madison Lu, Oisín …
7537  Alessandro Pepe  Leone Frisa, Paolo Vaccarino, Francesco Miglio…

        country date_added  release_year rating  duration  \
5989        NaN 2017-01-26          2017    NaN    37 min
6827      Japan 2016-12-01          2013    NaN  1 Season
7312  Australia 2018-02-01          2015    NaN  1 Season
7537      Italy 2017-03-01          2015    NaN   115 min

                               listed_in  \
5989                             Movies
6827  Anime Series, International TV Shows
7312                Kids' TV, TV Comedies
7537                             Dramas

                                      description
5989  Oprah Winfrey sits down with director Ava DuVe…
6827  After falling through a wormhole, a space-dwel…
7312  Adopting a child's perspective, this show take…
7537  Amid the chaos and horror of World War II, a c…
```

[ ]:
```
netflix.loc[netflix_NaN_index,'rating']="Not Available"
```

[ ]:
```
netflix.loc[netflix_NaN_index]
```

[ ]:
```
     show_id     type                                                 title  \
5989   s5990    Movie  13TH: A Conversation with Oprah Winfrey & Ava …
6827   s6828  TV Show                  Gargantia on the Verdurous Planet
```

```
7312   s7313  TV Show                                        Little Lunch
7537   s7538    Movie                                  My Honor Was Loyalty

             director                                                  cast  \
5989              NaN                          Oprah Winfrey, Ava DuVernay
6827              NaN  Kaito Ishikawa, Hisako Kanemoto, Ai Kayano, Ka…
7312              NaN  Flynn Curry, Olivia Deeble, Madison Lu, Oisín …
7537  Alessandro Pepe  Leone Frisa, Paolo Vaccarino, Francesco Miglio…

        country date_added  release_year         rating  duration  \
5989        NaN 2017-01-26          2017  Not Available    37 min
6827      Japan 2016-12-01          2013  Not Available  1 Season
7312  Australia 2018-02-01          2015  Not Available  1 Season
7537      Italy 2017-03-01          2015  Not Available   115 min

                                listed_in  \
5989                               Movies
6827  Anime Series, International TV Shows
7312                 Kids' TV, TV Comedies
7537                               Dramas

                                                 description
5989  Oprah Winfrey sits down with director Ava DuVe…
6827  After falling through a wormhole, a space-dwel…
7312  Adopting a child's perspective, this show take…
7537  Amid the chaos and horror of World War II, a c…
```

[ ]:
```
netflix['rating'].unique()
```

[ ]:
```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR',
       'Not Available', 'TV-Y7-FV', 'UR'], dtype=object)
```

Here we can see that rating has three unique values as minutes which should be in the duration field.

2.3)Cleaning Duration columns

11 4.3) Cleaning Duration Column

[ ]:
```
netflix[netflix['duration'].isna()]
```

[ ]:
```
     show_id   type                                    title    director  \
5541   s5542  Movie                      Louis C.K. 2017  Louis C.K.
5794   s5795  Movie                  Louis C.K.: Hilarious  Louis C.K.
5813   s5814  Movie  Louis C.K.: Live at the Comedy Store  Louis C.K.
```

```
            cast        country date_added  release_year  rating duration  \
5541  Louis C.K.  United States 2017-04-04          2017  74 min      NaN
5794  Louis C.K.  United States 2016-09-16          2010  84 min      NaN
5813  Louis C.K.  United States 2016-08-15          2015  66 min      NaN

     listed_in                                          description
5541    Movies  Louis C.K. muses on religion, eternal love, gi…
5794    Movies  Emmy-winning comedy writer Louis C.K. brings h…
5813    Movies  The comic puts his trademark hilarious/thought…
```

We have to update the values from rating to duration and also update the rating to Not available

[ ]:
```
netflix_NaN_duration=netflix[netflix['duration'].isna()].index
```

[ ]:
```
netflix.loc[netflix_NaN_duration,'duration']=netflix.
↪loc[netflix_NaN_duration,'rating']
```

[ ]:
```
netflix.loc[netflix_NaN_duration]
```

[ ]:
```
     show_id   type                                  title    director  \
5541   s5542  Movie                       Louis C.K. 2017  Louis C.K.
5794   s5795  Movie                   Louis C.K.: Hilarious  Louis C.K.
5813   s5814  Movie  Louis C.K.: Live at the Comedy Store  Louis C.K.

            cast        country date_added  release_year  rating duration  \
5541  Louis C.K.  United States 2017-04-04          2017  74 min   74 min
5794  Louis C.K.  United States 2016-09-16          2010  84 min   84 min
5813  Louis C.K.  United States 2016-08-15          2015  66 min   66 min

     listed_in                                          description
5541    Movies  Louis C.K. muses on religion, eternal love, gi…
5794    Movies  Emmy-winning comedy writer Louis C.K. brings h…
5813    Movies  The comic puts his trademark hilarious/thought…
```

Duration is updated

[ ]:
```
netflix.loc[netflix_NaN_duration,'rating']="Not Available"
```

[ ]:
```
netflix.loc[netflix_NaN_duration]
```

[ ]:
```
     show_id   type                                  title    director  \
5541   s5542  Movie                       Louis C.K. 2017  Louis C.K.
5794   s5795  Movie                   Louis C.K.: Hilarious  Louis C.K.
5813   s5814  Movie  Louis C.K.: Live at the Comedy Store  Louis C.K.

            cast        country date_added  release_year        rating  \
5541  Louis C.K.  United States 2017-04-04          2017  Not Available
5794  Louis C.K.  United States 2016-09-16          2010  Not Available
```

```
5813  Louis C.K.  United States 2016-08-15          2015  Not Available

     duration listed_in                                          description
5541   74 min    Movies  Louis C.K. muses on religion, eternal love, gi…
5794   84 min    Movies  Emmy-winning comedy writer Louis C.K. brings h…
5813   66 min    Movies  The comic puts his trademark hilarious/thought…
```

[ ]:
```
netflix['rating'].unique()
```

[ ]:
```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', 'Not Available', 'NR', 'TV-Y7-FV', 'UR'],
      dtype=object)
```

Rating field is updated

2.4) Cleaning director,cast and country

[ ]:
```
netflix['director'] = netflix['director'].fillna('Not Available')
netflix['cast'] = netflix['cast'].fillna('Not Available')
netflix['country'] = netflix['country'].fillna('Not Available')
```

Lets check the null values again

[ ]:
```
netflix.isna().sum()
```

[ ]:
```
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
```

Our dataset is cleaned up now

12 5) Exploring and Analysis of Data

[ ]:
```
netflix['type'].unique()
```

[ ]:
```
array(['Movie', 'TV Show'], dtype=object)
```

We can see that netflix has two types of videos. 1) Movies 2) TV Shows Let's make them two datasets to do better analysis

[ ]:
```
movies=netflix.loc[netflix['type']=='Movie'].copy()
movies.head()
```

[ ]:
```
   show_id   type                              title  \
0       s1  Movie              Dick Johnson Is Dead
6       s7  Movie  My Little Pony: A New Generation
7       s8  Movie                            Sankofa
9      s10  Movie                       The Starling
12     s13  Movie                       Je Suis Karl

                        director  \
0                Kirsten Johnson
6   Robert Cullen, José Luis Ucha
7                   Haile Gerima
9                 Theodore Melfi
12           Christian Schwochow

                                                 cast  \
0                                       Not Available
6   Vanessa Hudgens, Kimiko Glenn, James Marsden, …
7   Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D…
9   Melissa McCarthy, Chris O'Dowd, Kevin Kline, T…
12  Luna Wedler, Jannis Niewöhner, Milan Peschel, …

                                       country date_added  \
0                                United States 2021-09-25
6                                Not Available 2021-09-24
7   United States, Ghana, Burkina Faso, United Kin… 2021-09-24
9                                United States 2021-09-24
12                         Germany, Czech Republic 2021-09-23

    release_year rating duration  \
0           2020  PG-13   90 min
6           2021     PG   91 min
7           1993  TV-MA  125 min
9           2021  PG-13  104 min
12          2021  TV-MA  127 min

                                    listed_in  \
0                                Documentaries
6                     Children & Family Movies
7   Dramas, Independent Movies, International Movies
9                           Comedies, Dramas
12                Dramas, International Movies
```

```
                                                description
0    As her father nears the end of his life, filmm…
6    Equestria's divided. But a bright-eyed hero be…
7    On a photo shoot in Ghana, an American model s…
9    A woman adjusting to life after a loss contend…
12   After most of her family is murdered in a terr…
```

[ ]:
```
tv_shows=netflix.loc[netflix['type']=='TV Show'].copy()
tv_shows.head()
```

[ ]:
```
  show_id     type                  title         director  \
1      s2  TV Show          Blood & Water    Not Available
2      s3  TV Show              Ganglands  Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans    Not Available
4      s5  TV Show           Kota Factory    Not Available
5      s6  TV Show          Midnight Mass    Mike Flanagan

                                                  cast        country  \
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…   South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…  Not Available
3                                   Not Available  Not Available
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…          India
5  Kate Siegel, Zach Gilford, Hamish Linklater, H…  Not Available

  date_added  release_year rating   duration  \
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
5 2021-09-24          2021  TV-MA   1 Season

                                        listed_in  \
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 …
5                TV Dramas, TV Horror, TV Mysteries

                                          description
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…
5  The arrival of a charismatic young priest brin…
```

Since movie and TV shows both have different format for duration, we can change duration for

movies as minutes & TV shows as seasons

```
[ ]: movies['duration'] = movies['duration'].str[:-3]
     movies['duration'] = movies['duration'].astype('float')
```

```
[ ]: tv_shows['duration'] = tv_shows.duration.str[:-7].apply(lambda x : x.strip())
     tv_shows['duration'] = tv_shows['duration'].astype('float')
```

```
[ ]: tv_shows.rename({'duration': 'duration(seasons)'} ,axis = 1 , inplace = True)
     movies.rename({'duration': 'duration(minutes)'} ,axis = 1 , inplace = True)
```

Let's find out the the when was first movie/TV Show added on netflix and when is the most recent movie added on netflix

```
[ ]: date_added_min=netflix['date_added'].min()
     date_added_max=netflix['date_added'].max()
```

```
[ ]: timeperiod=pd.DataFrame({'date_added':
       ↪[date_added_min,date_added_max]},index=['First Movie Added','Most Recent␣
       ↪Movie Added'])
     timeperiod
```

```
[ ]:                         date_added
     First Movie Added       2008-01-01
     Most Recent Movie Added 2021-09-25
```

```
[ ]: releaseperiod=pd.DataFrame({'release_year':[netflix['release_year'].
       ↪min(),netflix['release_year'].max()]},index=['First Movie/TV Show Released␣
       ↪Year','Most Recent Movie/TV Show Released Year'])
     releaseperiod
```

```
[ ]:                                        release_year
     First Movie/TV Show Released Year              1925
     Most Recent Movie/TV Show Released Year        2021
```

Lets check Which are different ratings available on Netflix in each type of content? Check the number of content released in each type.

```
[ ]: netflix.groupby(['type','rating'])['show_id'].count()
```

```
[ ]: type     rating
     Movie    G                41
              NC-17             3
              NR               75
              Not Available     5
              PG              287
              PG-13           490
              R               797
```

```
          TV-14            1427
          TV-G              126
          TV-MA            2062
          TV-PG             540
          TV-Y              131
          TV-Y7             139
          TV-Y7-FV            5
          UR                  3
TV Show   NR                  4
          Not Available       2
          R                   2
          TV-14             730
          TV-G               94
          TV-MA            1143
          TV-PG             321
          TV-Y              175
          TV-Y7             194
          TV-Y7-FV            1
Name: show_id, dtype: int64
```

Lets Check the distribution of movies produced in different countries

[ ]:
```
netflix['country'].value_counts()
```

[ ]:
```
country
United States                                2812
India                                         972
Not Available                                 830
United Kingdom                                418
Japan                                         244
                                             …
Romania, Bulgaria, Hungary                      1
Uruguay, Guatemala                              1
France, Senegal, Belgium                        1
Mexico, United States, Spain, Colombia          1
United Arab Emirates, Jordan                    1
Name: count, Length: 749, dtype: int64
```

We can see that many movies are produced in more than one country. To analyse movies produced in each country, we need to split them into different rows by using explode function

[ ]:
```
country_table=netflix[['show_id','type','country']].copy()
country_table = country_table.drop(country_table[country_table['country'] ==␣
 ↪'Not Available'].index)
country_table['country']=country_table['country'].apply(lambda x : x.split(','))
country_table=country_table.explode('country')
country_table
```

[ ]:
```
     show_id     type        country
0          s1    Movie  United States
1          s2  TV Show   South Africa
4          s5  TV Show          India
7          s8    Movie  United States
7          s8    Movie          Ghana
…       …       …              …
8801    s8802    Movie         Jordan
8802    s8803    Movie  United States
8804    s8805    Movie  United States
8805    s8806    Movie  United States
8806    s8807    Movie          India

[10010 rows x 3 columns]
```

[ ]:
```
# delete all rows with unnecessary spaces
country_table['country'] = country_table['country'].str.strip()
```

[ ]:
```
country_table.loc[country_table['country'] == '']
country_table = country_table.loc[country_table['country'] != '']
country_table['country'].nunique()
```

[ ]:
```
122
```

Total 122 countries

Lets find out total Movies and TV Shows produced in each country:

[ ]:
```
x = country_table.groupby(['country' , 'type'])['show_id'].count().reset_index()
x.pivot(index = ['country'] , columns = 'type' , values = 'show_id').
 ↪sort_values('Movie',ascending = False)
```

[ ]:
```
type             Movie  TV Show
country
United States   2752.0    932.0
India            962.0     84.0
United Kingdom   534.0    271.0
Canada           319.0    126.0
France           303.0     90.0
…                …       …
Azerbaijan         NaN      1.0
Belarus            NaN      1.0
Cuba               NaN      1.0
Cyprus             NaN      1.0
Puerto Rico        NaN      1.0

[122 rows x 2 columns]
```

Lets do the same for director and cast column

```
[ ]: director_table=netflix[['show_id','type','director']].copy()
director_table = director_table.drop(director_table[director_table['director']␣
 ↪== 'Not Available'].index)
director_table['director']=director_table['director'].apply(lambda x : x.
 ↪split(','))
director_table
```

```
[ ]:      show_id     type                          director
0          s1    Movie                  [Kirsten Johnson]
2          s3  TV Show                  [Julien Leclercq]
5          s6  TV Show                    [Mike Flanagan]
6          s7    Movie  [Robert Cullen,  José Luis Ucha]
7          s8    Movie                     [Haile Gerima]
…       …       …                                  …
8801    s8802    Movie                  [Majid Al Ansari]
8802    s8803    Movie                    [David Fincher]
8804    s8805    Movie                  [Ruben Fleischer]
8805    s8806    Movie                     [Peter Hewitt]
8806    s8807    Movie                      [Mozez Singh]

[6173 rows x 3 columns]
```

```
[ ]: director_table=director_table.explode('director')
director_table['director'] = director_table['director'].str.strip()
```

```
[ ]: director_table
```

```
[ ]:      show_id     type         director
0          s1    Movie  Kirsten Johnson
2          s3  TV Show  Julien Leclercq
5          s6  TV Show    Mike Flanagan
6          s7    Movie    Robert Cullen
6          s7    Movie   José Luis Ucha
…       …       …              …
8801    s8802    Movie  Majid Al Ansari
8802    s8803    Movie    David Fincher
8804    s8805    Movie  Ruben Fleischer
8805    s8806    Movie     Peter Hewitt
8806    s8807    Movie      Mozez Singh

[6978 rows x 3 columns]
```

```
[ ]: director_table['director'].nunique()
```

```
[ ]: 4993
```

There are total 4993 unique directors in the dataset.

Lets find out total Movies and TV Shows directed by each director

[ ]:
```
x = director_table.groupby(['director' , 'type'])['show_id'].count().
 ↪reset_index()
x.pivot(index= ['director'] , columns = 'type' , values = 'show_id').
 ↪sort_values('Movie' ,ascending = False)
```

[ ]:
```
type                 Movie  TV Show
director
Rajiv Chilaka         22.0      NaN
Jan Suter             21.0      NaN
Raúl Campos           19.0      NaN
Suhas Kadav           16.0      NaN
Marcus Raboy          15.0      1.0
…                     …         …
Vijay S. Bhanushali    NaN      1.0
Wouter Bouvijn         NaN      1.0
YC Tom Lee             NaN      1.0
Yasuhiro Irie          NaN      1.0
Yim Pilsung            NaN      1.0

[4993 rows x 2 columns]
```

[ ]:
```
cast_table = netflix[['show_id' , 'type' ,'cast']].copy()
cast_table=cast_table.drop(cast_table[cast_table['cast'] == 'Not Available'].
 ↪index)
cast_table['cast'] = cast_table['cast'].apply(lambda x : x.split(','))
cast_table = cast_table.explode('cast')
cast_table
```

[ ]:
```
     show_id     type                    cast
1         s2  TV Show              Ama Qamata
1         s2  TV Show              Khosi Ngema
1         s2  TV Show           Gail Mabalane
1         s2  TV Show          Thabang Molaba
1         s2  TV Show        Dillon Windvogel
…       …       …                       …
8806   s8807    Movie        Manish Chaudhary
8806   s8807    Movie            Meghna Malik
8806   s8807    Movie           Malkeet Rauni
8806   s8807    Movie          Anita Shabdish
8806   s8807    Movie   Chittaranjan Tripathy

[64057 rows x 3 columns]
```

[ ]:
```
cast_table['cast'] = cast_table['cast'].str.strip()
```

[ ]:
```
cast_table['cast'].nunique()
```

[ ]:
```
36403
```

Lets find out total Movies and TV Shows acted by each cast

[ ]:
```
x = cast_table.groupby(['cast' , 'type'])['show_id'].count().reset_index()
x.pivot(index = ['cast'] , columns = 'type' , values = 'show_id').
 ↪sort_values('Movie',ascending = False)
```

[ ]:
```
type               Movie  TV Show
cast
Anupam Kher         42.0      1.0
Shah Rukh Khan      35.0      NaN
Naseeruddin Shah    32.0      NaN
Om Puri             30.0      NaN
Akshay Kumar        30.0      NaN
…                   …       …
İpek Filiz Yazıcı    NaN      1.0
İsmail Filiz         NaN      1.0
Şafak Başkaya        NaN      1.0
Şehsuvar Aktaş       NaN      1.0
Şenay Gürler         NaN      1.0

[36403 rows x 2 columns]
```

[ ]:
```
genre_table = netflix[['show_id' , 'type' ,'listed_in']].copy()
genre_table['listed_in'] = genre_table['listed_in'].apply(lambda x : x.
 ↪split(','))
genre_table = genre_table.explode('listed_in')
genre_table
```

[ ]:
```
     show_id     type                 listed_in
0         s1    Movie             Documentaries
1         s2  TV Show    International TV Shows
1         s2  TV Show                TV Dramas
1         s2  TV Show              TV Mysteries
2         s3  TV Show            Crime TV Shows
…       …       …                        …
8805   s8806    Movie  Children & Family Movies
8805   s8806    Movie                  Comedies
8806   s8807    Movie                   Dramas
8806   s8807    Movie      International Movies
8806   s8807    Movie          Music & Musicals

[19303 rows x 3 columns]
```

[ ]:
```
genre_table['listed_in'] = genre_table['listed_in'].str.strip()
```

[ ]:
```
genre_table['listed_in'].nunique()
```

[ ]:
```
42
```

[ ]:
```
x = genre_table.groupby(['listed_in' , 'type'])['show_id'].count().reset_index()
x.pivot(index = 'listed_in' , columns = 'type' , values = 'show_id').
 ↪sort_index()
```

[ ]:

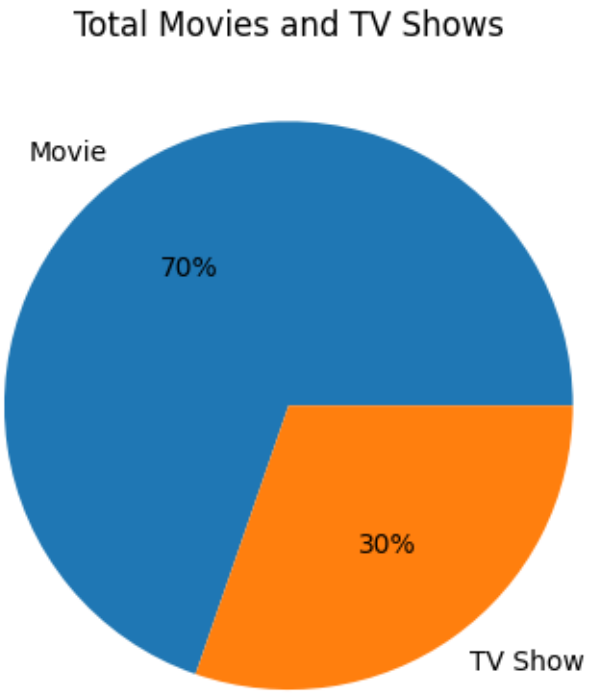
type	Movie	TV Show
listed_in		
Action & Adventure	859.0	NaN
Anime Features	71.0	NaN
Anime Series	NaN	175.0
British TV Shows	NaN	252.0
Children & Family Movies	641.0	NaN
Classic & Cult TV	NaN	26.0
Classic Movies	116.0	NaN
Comedies	1674.0	NaN
Crime TV Shows	NaN	469.0
Cult Movies	71.0	NaN
Documentaries	869.0	NaN
Docuseries	NaN	394.0
Dramas	2427.0	NaN
Faith & Spirituality	65.0	NaN
Horror Movies	357.0	NaN
Independent Movies	756.0	NaN
International Movies	2752.0	NaN
International TV Shows	NaN	1350.0
Kids' TV	NaN	449.0
Korean TV Shows	NaN	151.0
LGBTQ Movies	102.0	NaN
Movies	57.0	NaN
Music & Musicals	375.0	NaN
Reality TV	NaN	255.0
Romantic Movies	616.0	NaN
Romantic TV Shows	NaN	370.0
Sci-Fi & Fantasy	243.0	NaN
Science & Nature TV	NaN	92.0
Spanish-Language TV Shows	NaN	173.0
Sports Movies	219.0	NaN
Stand-Up Comedy	343.0	NaN
Stand-Up Comedy & Talk Shows	NaN	56.0
TV Action & Adventure	NaN	167.0
TV Comedies	NaN	574.0
TV Dramas	NaN	762.0

```
TV Horror                      NaN     75.0
TV Mysteries                   NaN     98.0
TV Sci-Fi & Fantasy            NaN     83.0
TV Shows                       NaN     16.0
TV Thrillers                   NaN     57.0
Teen TV Shows                  NaN     69.0
Thrillers                    577.0      NaN
```

13 6) Visual Analysis

Let's check the distribution of content across types

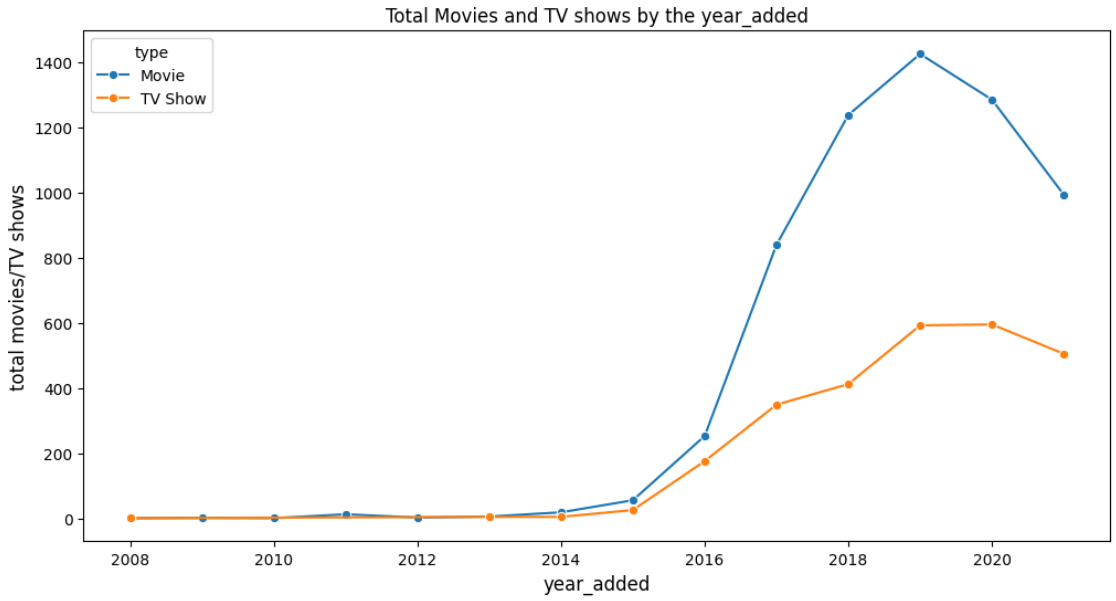
[ ]:
```
types=netflix['type'].value_counts()
plt.pie(types,  labels=types.index, autopct='%1.0f%%')
plt.title('Total Movies and TV Shows')
plt.show()
```



70% of Netflix content is Movies and 30% of content is TV Shows

[ ]:
```
netflix['year_added'] = netflix['date_added'].dt.year
```

[ ]:
```
# How has the number of movies/TV shows added on Netflix per year changed over␣
 ↪the time?
d = netflix.groupby(['year_added' ,'type' ])['show_id'].count().reset_index()
d.rename({'show_id' : 'total movies/TV shows'}, axis = 1 , inplace = True)
plt.figure(figsize = (12,6))
sns.lineplot(data = d , x = 'year_added' , y = 'total movies/TV shows' , hue =␣
 ↪'type', marker = 'o'  , ms = 6)
plt.xlabel('year_added' , fontsize = 12)
plt.ylabel('total movies/TV shows' , fontsize = 12)
plt.title('Total Movies and TV shows by the year_added' , fontsize = 12)
plt.show()
```



The content added on the Netflix surged drastically after 2015.

2019 marks the highest number of movies and TV shows added on the Netflix.

Year 2020 and 2021 has seen the drop in content added on Netflix, possibly because of Pandemic.

But still , TV shows content have not dropped as drastic as movies. In recent years TV shows are focussed more than Movies.

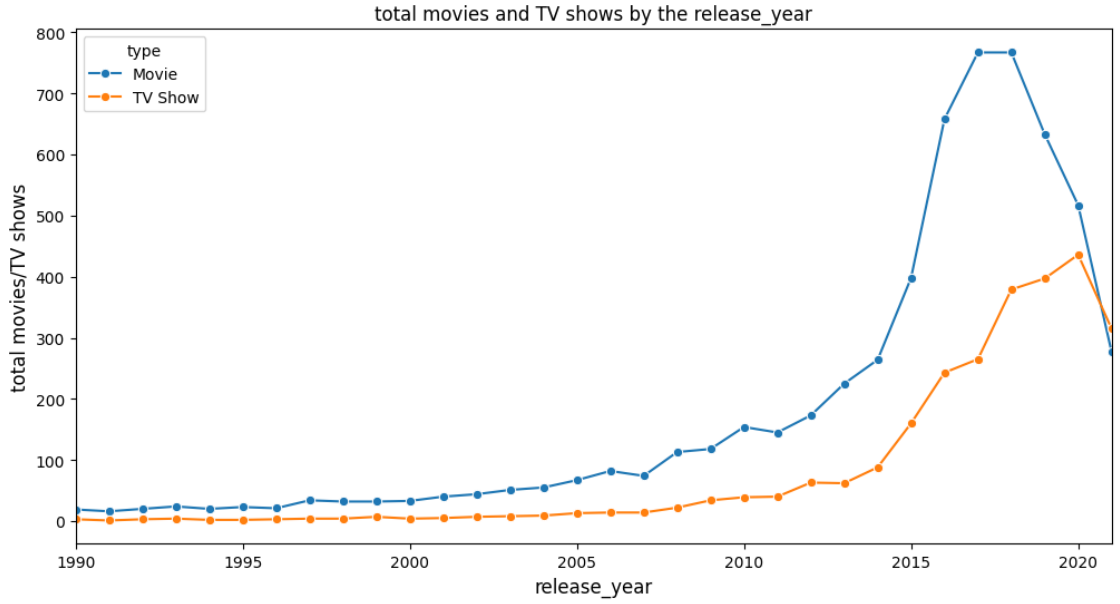
[ ]:
```
d = netflix.groupby(['type' , 'release_year'])['show_id'].count().reset_index()
d.rename({'show_id' : 'total movies/TV shows'}, axis = 1 , inplace = True)
d
```

[ ]:

	type	release_year	total movies/TV shows
0	Movie	1942	2
1	Movie	1943	3
2	Movie	1944	3

```
3       Movie          1945                     3
4       Movie          1946                     1
..        …            …                      …
114   TV Show          2017                   265
115   TV Show          2018                   379
116   TV Show          2019                   397
117   TV Show          2020                   436
118   TV Show          2021                   315

[119 rows x 3 columns]
```

[ ]:
```
plt.figure(figsize = (12,6))
sns.lineplot(data = d , x = 'release_year' , y = 'total movies/TV shows' , hue␣
 ↪= 'type' , marker = 'o'  , ms = 6 )
plt.xlabel('release_year' , fontsize = 12)
plt.ylabel('total movies/TV shows' , fontsize = 12)
plt.title('total movies and TV shows by the release_year' , fontsize = 12)
plt.xlim( left = 1990 , right = 2021)
plt.show()
```



2018 marks the highest number of movie and TV show releases.

Since 2018, A drop in movies is seen and rise in TV shows is observed clearly, and TV shows surpasses the movies count in mid 2020.

In recent years TV shows are focussed more than Movies.

The yearly number of releases has surged drastically from 2015.

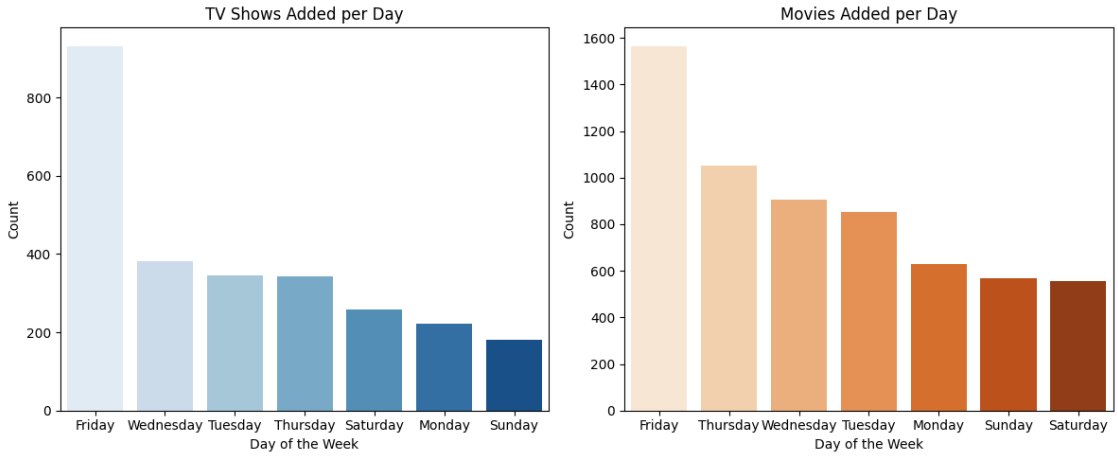
[ ]:
```
tv_shows.head()
```

[ ]:
```
  show_id     type                  title         director  \
1      s2  TV Show          Blood & Water    Not Available
2      s3  TV Show              Ganglands  Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans    Not Available
4      s5  TV Show           Kota Factory    Not Available
5      s6  TV Show          Midnight Mass    Mike Flanagan

                                                  cast        country  \
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…   South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi…  Not Available
3                                    Not Available  Not Available
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K…          India
5  Kate Siegel, Zach Gilford, Hamish Linklater, H…  Not Available

  date_added  release_year rating  duration(seasons)  \
1 2021-09-24          2021  TV-MA                2.0
2 2021-09-24          2021  TV-MA                1.0
3 2021-09-24          2021  TV-MA                1.0
4 2021-09-24          2021  TV-MA                2.0
5 2021-09-24          2021  TV-MA                1.0

                                        listed_in  \
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 …
5                TV Dramas, TV Horror, TV Mysteries

                                        description
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…
5  The arrival of a charismatic young priest brin…
```

[ ]:
```
tv_shows['day']=pd.to_datetime(tv_shows['date_added']).dt.day_name()
movies['day']=pd.to_datetime(movies['date_added']).dt.day_name()
```

[ ]:
```
tv_counts = tv_shows['day'].value_counts()
movies_counts = movies['day'].value_counts()
country=tv_shows['country'].value_counts()

# Create subplots
fig, ax = plt.subplots(1, 2, figsize=(12, 5))
```

```
# TV Shows subplot
sns.barplot(x=tv_counts.index, y=tv_counts.values, ax=ax[0], hue=tv_counts.
 ↪index, palette='Blues', legend=False)
ax[0].set_title("TV Shows Added per Day")
ax[0].set_xlabel("Day of the Week")
ax[0].set_ylabel("Count")

# Movies subplot
sns.barplot(x=movies_counts.index, y=movies_counts.values, ax=ax[1],␣
 ↪hue=movies_counts.index, palette='Oranges', legend=False)
ax[1].set_title("Movies Added per Day")
ax[1].set_xlabel("Day of the Week")
ax[1].set_ylabel("Count")

plt.tight_layout()
plt.show()
```



It is seen that most of the TV Shows and Movies are released on Fridays

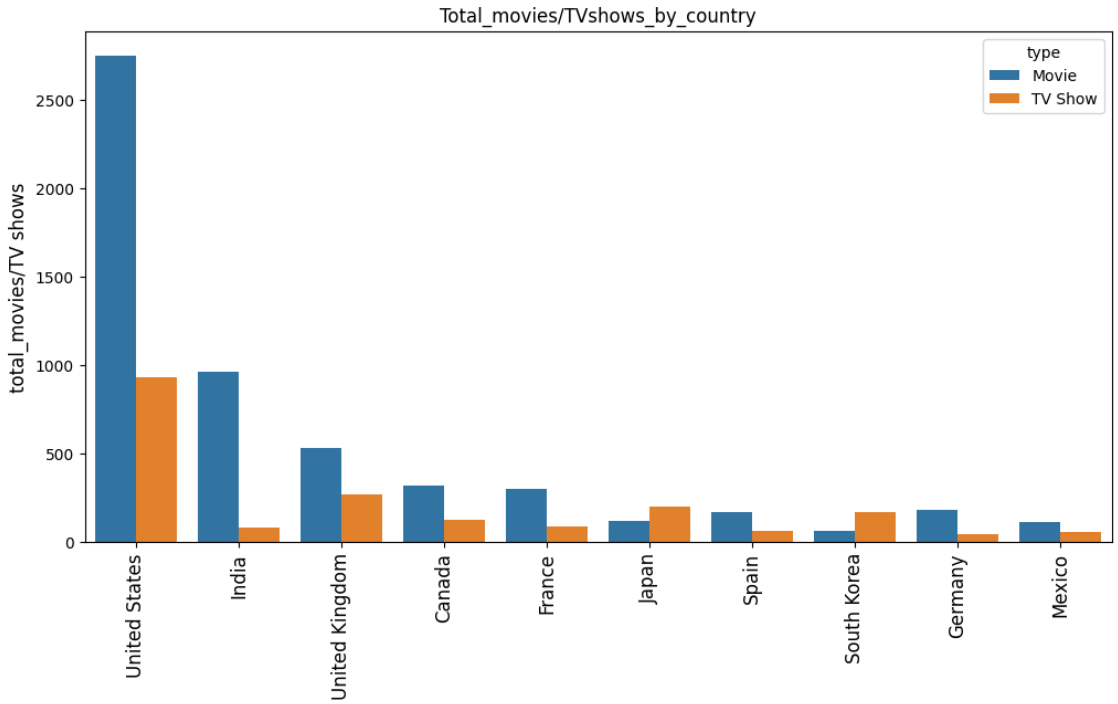
[ ]:
```
# Lets check for top 10 countries
top_10_country = country_table['country'].value_counts().head(10).index
df_new = country_table.loc[country_table['country'].isin(top_10_country)]
```

[ ]:
```
x = df_new.groupby(['country' , 'type'])['show_id'].count().reset_index()
x.pivot(index = 'country' , columns = 'type' , values = 'show_id').
 ↪sort_values('Movie',ascending = False)
```

[ ]:
```
type            Movie  TV Show
country
United States    2752      932
India             962       84
```

```
United Kingdom    534      271
Canada            319      126
France            303       90
Germany           182       44
Spain             171       61
Japan             119      198
Mexico            111       58
South Korea        61      170
```

[ ]:
```
plt.figure(figsize= (12,6))
sns.countplot(data = df_new , x = 'country' , order = top_10_country , hue =␣
↪'type')
plt.xticks(rotation = 90 , fontsize = 12)
plt.ylabel('total_movies/TV shows' , fontsize = 12)
plt.xlabel('')
plt.title('Total_movies/TVshows_by_country')
plt.show()
```

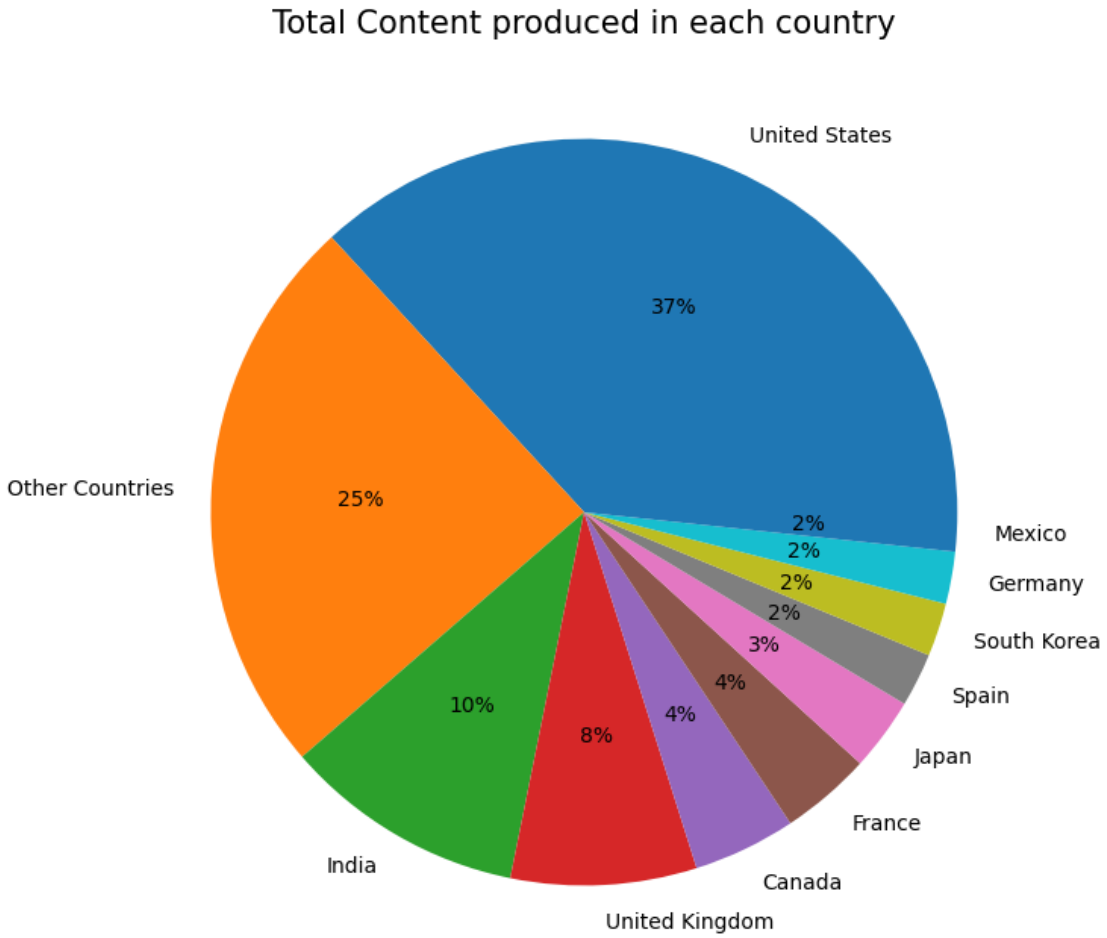


United States is leading the table with most contents on Netflix

[83]:
```
# Get the top 10 countries
top_10_country = country_table['country'].value_counts().head(10).index

# Explicitly create a copy of the DataFrame
```

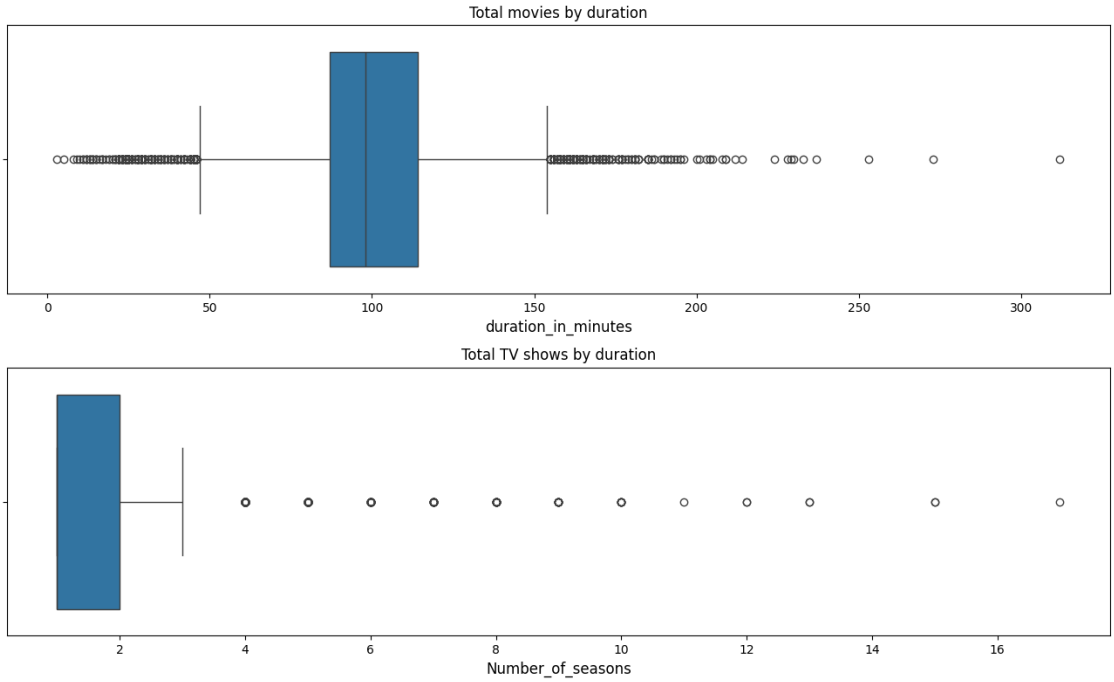
```
country_table = country_table.copy()

# Create a new column 'cat' and assign values
country_table['cat'] = country_table['country'].apply(lambda x: x if x in␣
 ↪top_10_country else 'Other Countries')
```

[84]:
```
x = country_table['cat'].value_counts()

plt.figure(figsize = (8,8))
plt.pie(x , labels = x.index, autopct='%1.0f%%')
plt.title('Total Content produced in each country' , fontsize = 15)
plt.show()
```



Movies and TV Shows Distribution over duration

[ ]:
```
fig, ax = plt.subplots(2,1, figsize=(13,8))

sns.boxplot (data = movies , x = 'duration(minutes)' ,ax =ax[0])
ax[0].set_xlabel('duration_in_minutes' ,  fontsize = 12)
ax[0].set_title('Total movies by duration')

sns.boxplot (data = tv_shows , x = 'duration(seasons)' , ax = ax[1])
ax[1].set_xlabel('Number_of_seasons' ,  fontsize = 12)
ax[1].set_title('Total TV shows by duration')

plt.tight_layout()
plt.show()
```

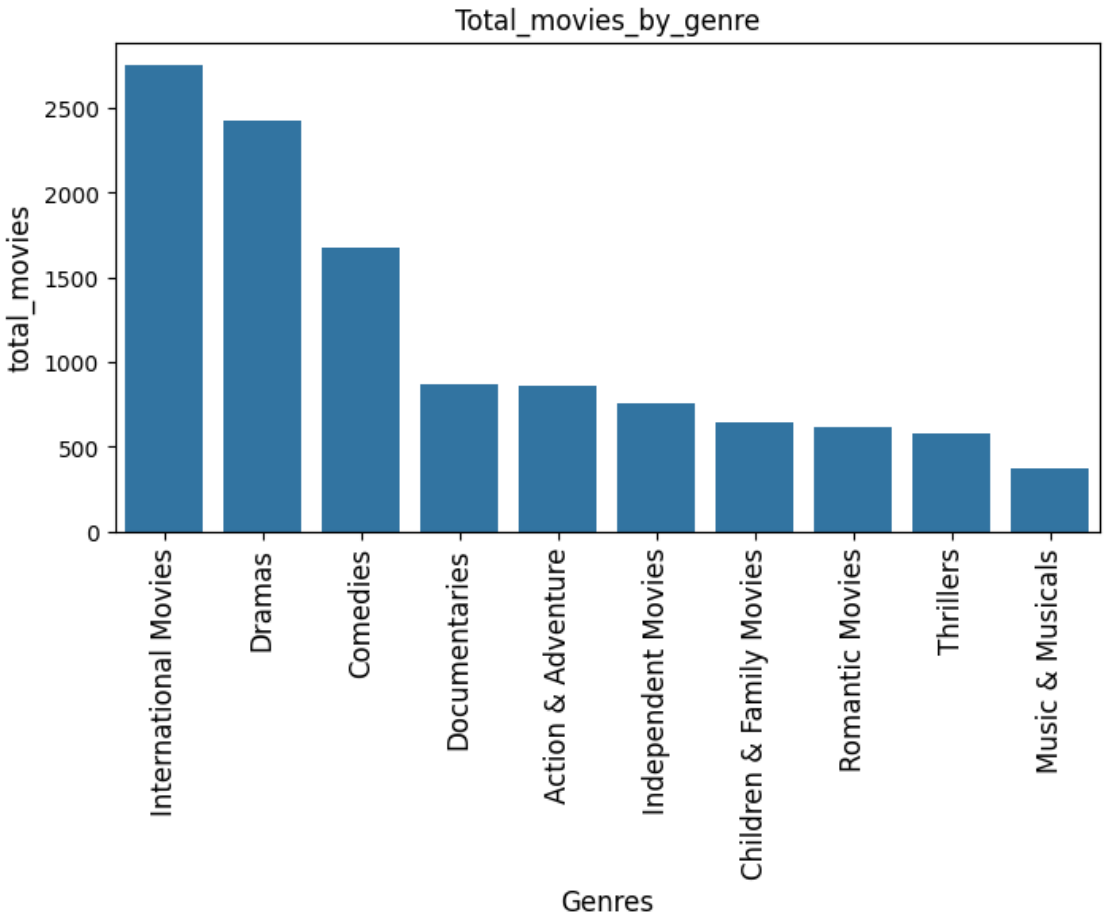


- **Movie Duration: 50 mins - 150 mins is the range excluding potential outliers (values lying outside the whiskers of boxplot)**
- **TV Show Duration: 1-3 seasons is the range for TV shows excluding potential outliers**

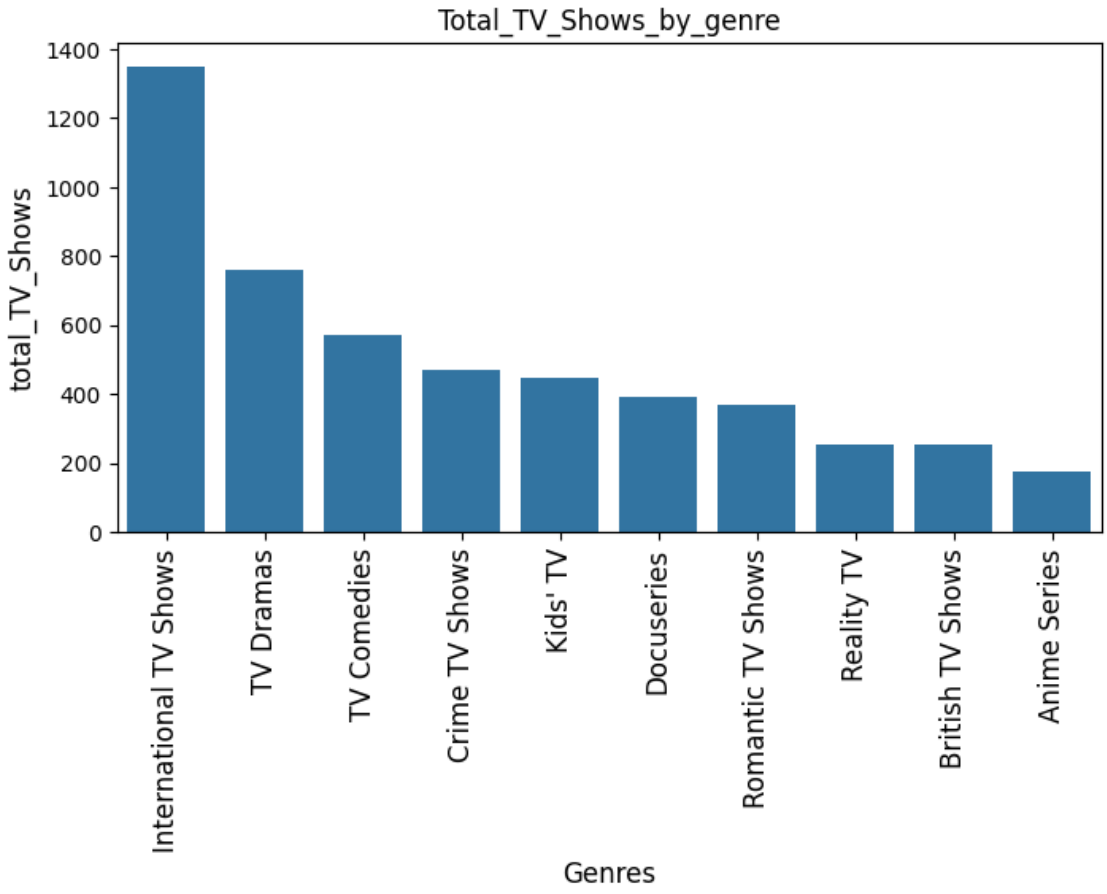
[ ]:
```
top_10_movie_genres = genre_table[genre_table['type'] == 'Movie'].listed_in.
 ↪value_counts().head(10).index
df_movie = genre_table.loc[genre_table['listed_in'].isin(top_10_movie_genres)]
```

[ ]:
```
top_10_TV_genres = genre_table[genre_table['type'] == 'TV Show'].listed_in.
 ↪value_counts().head(10).index
df_tv = genre_table.loc[genre_table['listed_in'].isin(top_10_TV_genres)]
```

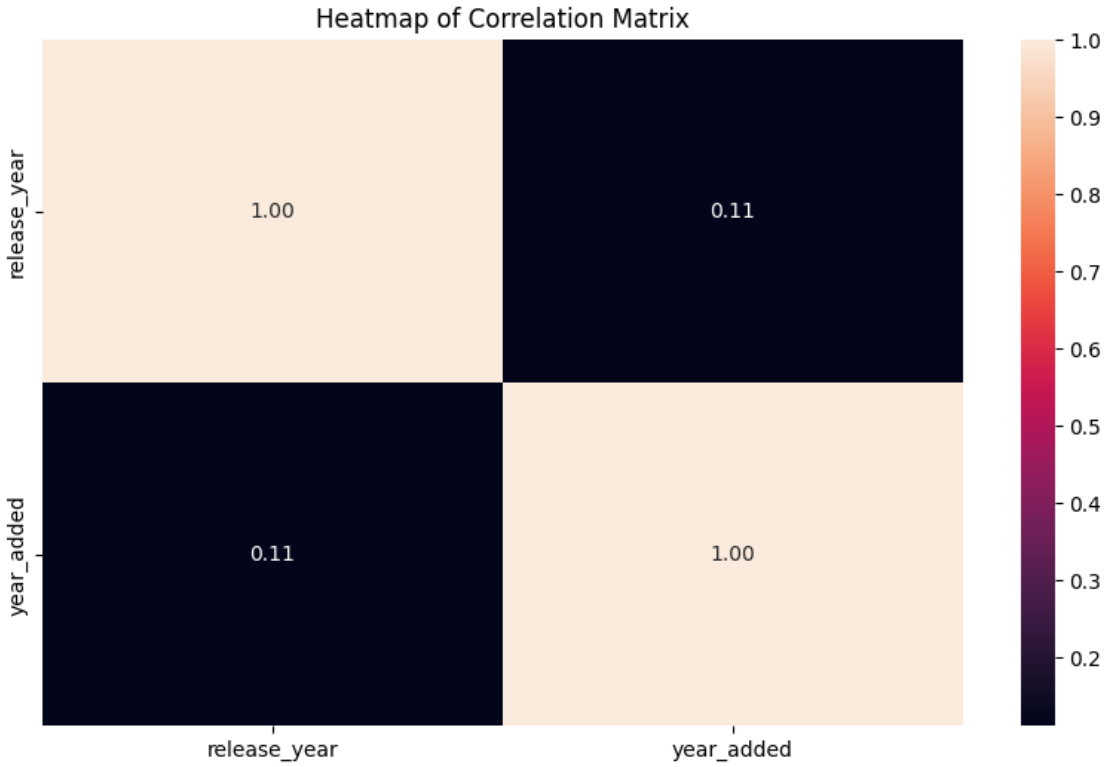
[ ]:
```
plt.figure(figsize= (8,4))
sns.countplot(data = df_movie , x = 'listed_in' , order = top_10_movie_genres)
plt.xticks(rotation = 90 , fontsize = 12)
plt.ylabel('total_movies' , fontsize = 12)
plt.xlabel('Genres' , fontsize = 12)
plt.title('Total_movies_by_genre')
plt.show()
```



[ ]:
```
plt.figure(figsize= (8,4))
sns.countplot(data = df_tv , x = 'listed_in' ,order = top_10_TV_genres)
plt.xticks(rotation = 90 , fontsize = 12)
plt.ylabel('total_TV_Shows' , fontsize = 12)
plt.xlabel('Genres' , fontsize = 12)
plt.title('Total_TV_Shows_by_genre')
plt.show()
```



[ ]:
```
df_numeric = netflix.select_dtypes(include=['number'])
corr_matrix = df_numeric.corr()

plt.figure(figsize=(10, 6))
sns.heatmap(corr_matrix, annot=True, fmt=".2f")
plt.title("Heatmap of Correlation Matrix")
plt.show()
```



Weak Correlation (0.11) Between Release Year and Year Added:

The low correlation (0.11) suggests that there is no strong relationship between when a movie/TV show was released and when it was added to the platform.

Content is Not Immediately Added After Release:

Since the correlation is weak, Netflix does not necessarily add content close to its release year. Instead, older content is also frequently added.

Netflix's Licensing and Acquisition Strategy is Diverse:

The low correlation indicates that Netflix adds content based on factors other than just release year, such as licensing agreements, audience demand, and strategic partnerships.

14 7)Outlier Treatment

[ ]:
```
num_cols = ['release_year']
```

[ ]:
```
df = netflix.dropna(subset=num_cols)
```

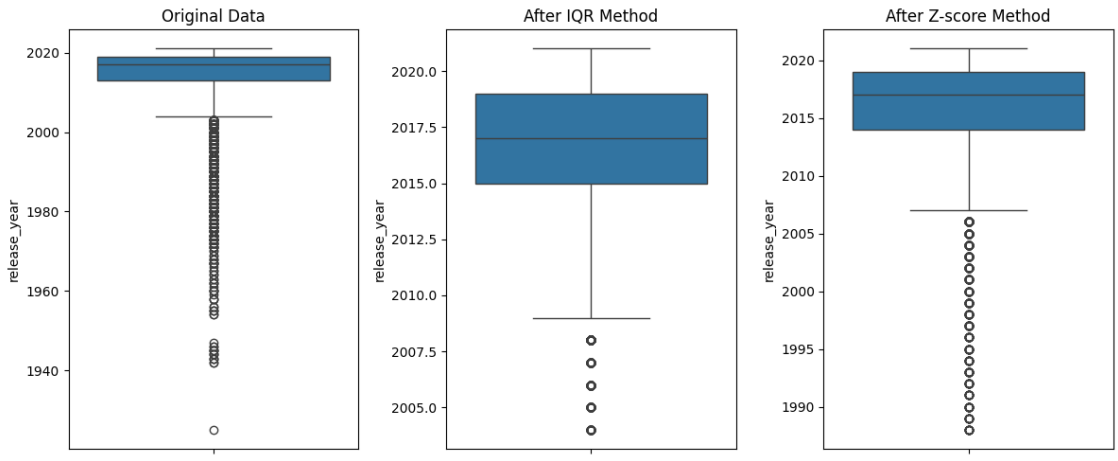
[ ]:
```
def remove_outliers_iqr(data, column):
    Q1 = data[column].quantile(0.25)
    Q3 = data[column].quantile(0.75)
```

```
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    return data[(data[column] >= lower_bound) & (data[column] <= upper_bound)]
```

[ ]:
```
def remove_outliers_zscore(data, column, threshold=3):
    mean = data[column].mean()
    std = data[column].std()
    return data[np.abs((data[column] - mean) / std) < threshold]
```

[ ]:
```
df_iqr = remove_outliers_iqr(netflix, 'release_year')
df_zscore = remove_outliers_zscore(netflix, 'release_year')
```

[ ]:
```
plt.figure(figsize=(12, 5))

plt.subplot(1, 3, 1)
sns.boxplot(y=netflix['release_year'])
plt.title("Original Data")

plt.subplot(1, 3, 2)
sns.boxplot(y=df_iqr['release_year'])
plt.title("After IQR Method")

plt.subplot(1, 3, 3)
sns.boxplot(y=df_zscore['release_year'])
plt.title("After Z-score Method")

plt.tight_layout()
plt.show()
```



IQR is better for skewed distributions since it focuses on the middle 50% of data.

Z-score works well when data is normally distributed, as it relies on standard deviation.

Final choice depends on business needs—IQR is robust, while Z-score can be useful for symmetrical data

15 Insights based on Non-Graphical and Visual Analysis

1. Movies dominate Netflix content (~70%), while TV shows make up ~30%. Since 2015, TV shows have surged, surpassing movies in mid-2020.
2. Content growth peaked in 2019, with a decline in 2020-21, likely due to the pandemic. TV shows remained more stable compared to movies.
3. Diverse global content—Netflix hosts movies/shows from 122 countries, with the U.S. contributing 37%. India has the most actors featured.
4. Content ratings vary by region—TV-MA is the most common, while kid-friendly ratings are more prevalent in the U.S.
5. Top genres include International Movies/TV Shows, Dramas, and Comedies, with region-specific trends (e.g., Korean TV Shows in Korea, Anime in Japan).
6. Viewing preferences have evolved—Recent years show a preference for shorter-duration movies (50-150 mins) and TV shows with 1-3 seasons.

16 Business Insights

1. **Limited Pre-2000 Content**: The majority of Netflix's catalog consists of post-2000 releases. Older content is scarce, presenting an opportunity to target senior citizens with classic films and shows.
2. **Focus on Mature Audiences** : Over 80% of Netflix content is rated TV-MA, TV-14, TV-PG, or R, catering primarily to mature viewers and children with parental guidance. Expanding kid-friendly or all-age content could attract a broader audience.
3. **Genre Trends**: International Movies/TV Shows, Dramas, Comedies, Action & Adventure, and Thrillers are the most popular genres, highlighting strong global and entertainment-driven content preferences.
4. **Content Source Concentration**: 75% of Netflix content comes from just 10 countries, while the rest of the world contributes only 25%. Expanding content from underrepresented regions could drive growth and diversification.
5. **Preference for Shorter Content** : Viewers favor movies between 75-150 minutes and TV shows with 1-3 seasons. New content production should align with this trend to maximize engagement.
6. **Pandemic Impact on Content**: A noticeable drop in content additions in 2020-2021 across all countries, likely due to the pandemic, suggests a temporary slowdown in production and acquisitions.

17 Recommendations

1. **Shift from Movies to TV Shows** : Netflix initially had 2.3x more movies than TV shows (6131 vs. 2676). However, in the past five years, the movie proportion has dropped from ~75% to ~47%, while TV shows have surged from ~25% to ~53%. This shift indicates a growing demand for episodic content over standalone films.

2. **Gaps in Audience Targeting** : Around 80% of Netflix content caters to mature audiences (TV-MA, TV-14, TV-PG) or children with parental guidance. Expanding content for families, senior citizens, and kids of various age groups could unlock new audience segments.

3. **Growing Preference for Shorter Content** : In the last decade, viewers have shown a clear preference for movies lasting 75-150 minutes and TV series with 1-3 seasons. Aligning new productions with this trend could enhance engagement.

4. **Scarcity of Pre-2000 Content** : Most of Netflix's library consists of content released after 2000, with very little older content available. Senior citizens, who may seek classic films and TV shows, are currently an underserved audience.

5. **Need for Regional Genre Expansion** : Netflix content is heavily US-centric, with limited genre variety in most countries. Introducing country-specific genres (e.g., Indian Mythological content or Japanese Anime) could drive local and global interest, just as anime has gained worldwide popularity.

6. ** Country-Specific Content Strategies** : Content distribution should align with a country's demographic needs.

For example:

India : Mostly serves TV-MA, TV-14, and TV-PG content, which may not cater to kids under 14 and adults over 35.

Japan : Serves mainly TV-MA, TV-14, and TV-PG, despite having a large aging population (60+ years). Increasing age-appropriate content could better serve this demographic.

7. **Content Source Concentration** : 75% of Netflix's library comes from just 10 countries, with the United States, India, and the UK as top contributors. Expanding content from underrepresented regions with relevant genres and ratings could fuel global growth.

#The link to the colab file

https://colab.research.google.com/drive/1iSQQNaeqo5zandvpRY6TgwOW2_PVb8Px?usp=sharing

[ ]: