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. 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.

Business Problem ¶

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

Dataset

The dataset provided to you consists of a list of all the TV shows/movies available on Netflix:

- Show_id: Unique ID for every Movie / Tv Show
- Type: Identifier - A Movie or TV Show
- Title: Title of the Movie / Tv Show
- Director: Director of the Movie
- Cast: Actors involved in the movie/show
- Country: Country where the movie/show was produced
- Date_added: Date it was added on Netflix
- Release_year: Actual Release year of the movie/show
- Rating: TV Rating of the movie/show
- Duration: Total Duration - in minutes or number of seasons
- Listed_in: Genre
- Description: The summary description

In [ ]:

1. Defining Problem Statement and Analysing basic metrics

Import Libraries

Importing the libraries we need

In [1]:

```
import mysql.connector
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
# Establishing the connection
db_connection = mysql.connector.connect(
    host='localhost',
    user='root',
    password='root',
    database='netflix'
)

# Creating a cursor object
cursor = db_connection.cursor()

# Executing a query
cursor.execute("SELECT *from netflix_shows")
# Fetch all data
netflix_df = pd.DataFrame(cursor.fetchall(), columns=[i[0] for i in cursor.
description])
netflix_df
```

Out[1]:

	show_id	type	title	director	cast	country	date_added	release_ye
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	None	United States	September 25, 2021	20
1	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	20
2	s11	TV Show	Vendetta: Truth, Lies and The Mafia	None	None	None	September 24, 2021	20
3	s12	TV Show	Bangkok Breaking	Kongkiat Komesiri	Sukollawat Kanarot, Sushar Manaying, Pavarit M...	None	September 23, 2021	20
4	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	20
5	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Klara Castanho, Lucca Picon, Júlia Gomes, Marc...	None	September 22, 2021	20
6	s15	TV Show	Crime Stories: India Detectives	None	None	None	September 22, 2021	20
7	s16	TV Show	Dear White People	None	Logan Browning, Brandon P. Bell, DeRon Horton,...	United States	September 22, 2021	20
8	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	Pedro de Echave García, Pablo Azorín Williams	None	None	September 22, 2021	20
9	s18	TV Show	Falsa identidad	None	Luis Ernesto Franco, Camila Sodi, Sergio Goyri...	Mexico	September 22, 2021	20
10	s19	Movie	Intrusion	Adam Salky	Freida Pinto, Logan Marshall-Green, Robert Joh...	None	September 22, 2021	20

	show_id	type	title	director	cast	country	date_added	release_ye
11	s2	TV Show	Blood & Water	None	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	20
12	s20	TV Show	Jaguar	None	Blanca Suárez, Iván Marcos, Óscar Casas, Adriá...	None	September 22, 2021	20
13	s21	TV Show	Monsters Inside: The 24 Faces of Billy Milligan	Olivier Megaton	None	None	September 22, 2021	20
14	s22	TV Show	Resurrection: Ertugrul	None	Engin Altan Düzyatan, Serdar Gökhan, Hülya Dar...	Turkey	September 22, 2021	20
15	s23	Movie	Avvai Shanmughi	K.S. Ravikumar	Kamal Hassan, Meena, Gemini Ganesan, Heera Raj...	None	September 21, 2021	19
16	s24	Movie	Go! Go! Cory Carson: Chrissy Takes the Wheel	Alex Woo, Stanley Moore	Maisie Benson, Paul Killam, Kerry Gudjohnsen, ...	None	September 21, 2021	20
17	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	September 21, 2021	19
18	s26	TV Show	Love on the Spectrum	None	Brooke Satchwell	Australia	September 21, 2021	20
19	s27	Movie	Minsara Kanavu	Rajiv Menon	Arvind Swamy, Kajol, Prabhu Deva, Nassar, S.P....	None	September 21, 2021	19
20	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	September 20, 2021	20
21	s29	Movie	Dark Skies	Scott Stewart	Keri Russell, Josh Hamilton, J.K. Simmons, Dak...	United States	September 19, 2021	20

	show_id	type	title	director	cast	country	date_added	release_ye
22	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	None	September 24, 2021	20
23	s30	Movie	Paranoia	Robert Luketic	Liam Hemsworth, Gary Oldman, Amber Heard, Harr...	United States, India, France	September 19, 2021	20
24	s31	Movie	Ankahi Kahaniya	Ashwiny Iyer Tiwari, Abhishek Chaubey, Saket C...	Abhishek Banerjee, Rinku Rajguru, Delzad Hiwal...	None	September 17, 2021	20
25	s32	TV Show	Chicago Party Aunt	None	Lauren Ash, Rory O'Malley, RuPaul Charles, Jil...	None	September 17, 2021	20
26	s33	TV Show	Sex Education	None	Asa Butterfield, Gillian Anderson, Ncuti Gatwa...	United Kingdom	September 17, 2021	20
27	s34	TV Show	Squid Game	None	Lee Jung-jae, Park Hae-soo, Wi Ha-jun, Oh Youn...	None	September 17, 2021	20
28	s35	TV Show	Tayo and Little Wizards	None	Dami Lee, Jason Lee, Bommie Catherine Han, Jen...	None	September 17, 2021	20
29	s36	Movie	The Father Who Moves Mountains	Daniel Sandu	Adrian Titieni, Elena Purea, Judith State, Val...	None	September 17, 2021	20
30	s37	Movie	The Stronghold	Cédric Jimenez	Gilles Lellouche, Karim Leklou, François Civil...	None	September 17, 2021	20
31	s38	TV Show	Angry Birds	None	Antti Pääkkönen, Heljä Heikkinen, Lynne Guagli...	Finland	September 16, 2021	20
32	s39	Movie	Birth of the Dragon	George Nolfi	Billy Magnussen, Ron Yuan, Qu Jingjing, Terry ...	China, Canada, United States	September 16, 2021	20

	show_id	type	title	director	cast	country	date_added	release_ye
33	s4	TV Show	Jailbirds New Orleans	None	None	None	September 24, 2021	20
34	s40	TV Show	Chhota Bheem	None	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jig...	India	September 16, 2021	20
35	s41	TV Show	He-Man and the Masters of the Universe	None	Yuri Lowenthal, Kimberly Brooks, Antony Del Ri...	United States	September 16, 2021	20
36	s42	Movie	Jaws	Steven Spielberg	Roy Scheider, Robert Shaw, Richard Dreyfuss, L...	United States	September 16, 2021	19
37	s43	Movie	Jaws 2	Jeannot Szwarc	Roy Scheider, Lorraine Gary, Murray Hamilton, ...	United States	September 16, 2021	19
38	s44	Movie	Jaws 3	Joe Alves	Dennis Quaid, Bess Armstrong, Simon MacCorkind...	United States	September 16, 2021	19
39	s45	Movie	Jaws: The Revenge	Joseph Sargent	Lorraine Gary, Lance Guest, Mario Van Peebles,...	United States	September 16, 2021	19
40	s46	Movie	My Heroes Were Cowboys	Tyler Greco	None	None	September 16, 2021	20
41	s47	Movie	Safe House	Daniel Espinosa	Denzel Washington, Ryan Reynolds, Vera Farmiga...	South Africa, United States, Japan	September 16, 2021	20
42	s48	TV Show	The Smart Money Woman	Bunmi Ajakaiye	Osas Ighodaro, Ini Dima-Okojie, Kemi Lala Akin...	None	September 16, 2021	20
43	s49	Movie	Training Day	Antoine Fuqua	Denzel Washington, Ethan Hawke, Scott Glenn, T...	United States	September 16, 2021	20
44	s5	TV Show	Kota Factory	None	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	20

	show_id	type	title	director	cast	country	date_added	release_ye
45	s50	TV Show	Castle and Castle	None	Richard Mofe-Damijo, Dakore Akande, Bimbo Manu...	Nigeria	September 15, 2021	20
46	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, H...	None	September 24, 2021	20
47	s7	Movie	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	None	September 24, 2021	20
48	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	19
49	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021	20

Let’s check the first 5 data.

In [2]:
```
netflix_df.head()
```

Out[2]:

	show_id	type	title	director	cast	country	date_added	release_year	rat
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	None	United States	September 25, 2021	2020	P
1	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	P
2	s11	TV Show	Vendetta: Truth, Lies and The Mafia	None	None	None	September 24, 2021	2021	
3	s12	TV Show	Bangkok Breaking	Kongkiat Komesiri	Sukollawat Kanarot, Sushar Manaying, Pavarit M...	None	September 23, 2021	2021	
4	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	

2: Observations on the shape of data, data types of allthe attributes, conversion of categorical attributes to'category' (If required), missing value detection,statistical summary

In [3]:
```
# To get All attributes netflix_df.columns
netflix_df.columns
```

Out[3]:
```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_ad
ded',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

In [4]:
```
netflix_df.shape
```

Out[4]:
```
(50, 12)
```

In [5]:
```
# Data types of all the attributes
netflix_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       50 non-null     object
 1   type          50 non-null     object
 2   title         50 non-null     object
 3   director      32 non-null     object
 4   cast          43 non-null     object
 5   country       27 non-null     object
 6   date_added    50 non-null     object
 7   release_year  50 non-null     int64 
 8   rating        50 non-null     object
 9   duration      50 non-null     object
 10  listed_in     50 non-null     object
 11  description   50 non-null     object
dtypes: int64(1), object(11)
memory usage: 4.8+ KB
```

In [6]:
```
# Statistical Summary Before Data Cleaning:
netflix_df.describe()
```

Out[6]:

	release_year
count	50.000000
mean	2014.340000
std	12.479551
min	1975.000000
25%	2014.000000
50%	2021.000000
75%	2021.000000
max	2021.000000

Missing Value Detection

Data Profiling & Cleaning

Data Cleaning means the process of identifying incorrect, incomplete, inaccurate, irrelevant, or missing pieces of data and then modifying, replacing, or deleting them as needed. Data Cleansing is considered as the basic element of Data Science.

In [7]:
```
print('\nColumns with missing value:')
print(netflix_df.isnull().any())
```

```
Columns with missing value:
show_id         False
type            False
title           False
director         True
cast             True
country          True
date_added      False
release_year    False
rating          False
duration        False
listed_in       False
description     False
dtype: bool
```

From the info, we know that there are 50 entries and 12 columns to work with for this EDA. There are a few columns that contain null values, “director,” “cast,” “country,”.

In [8]:
```
netflix_df.T.apply(lambda x: x.isnull().sum(), axis = 1)
```

Out[8]:
```
show_id          0
type             0
title            0
director        18
cast             7
country         23
date_added       0
release_year     0
rating           0
duration         0
listed_in        0
description      0
dtype: int64
```

In [9]:
```
netflix_df.isnull().sum().sum()
```

Out[9]:
```
np.int64(48)
```

There are a total of 48 null values across the entire dataset with 18 under "director", 7 under "cast", 23 under "country". We will have to handle all null data points before

---

we can dive into EDA and modelling.

Imputation is a treatment method for missing value by filling it in using certain techniques.

Can use mean, mode, or use predictive modelling. In thiscase study, we will discuss the use of the fillna function from Pandas for this imputation. Drop rows containing missing values. Can use the dropna function from Pandas.

In [10]:
```
netflix_df['director'] = netflix_df['director'].fillna("No Director")
netflix_df['cast'] = netflix_df['cast'].fillna("No Cast")
netflix_df['country'] = netflix_df['country'].fillna("Country Unavailable")
netflix_df.dropna(subset=["duration", "rating"], inplace=True)
```

In [11]:
```
netflix_df.isna().any()
```

Out[11]:
```
show_id         False
type            False
title           False
director        False
cast            False
country         False
date_added      False
release_year    False
rating          False
duration        False
listed_in       False
description     False
dtype: bool
```

For missing values, the easiest way to get rid of them would be to delete the rows with the missing data. However, this wouldn't be beneficial to our EDA since the is a loss of information. Since "director", "cast", and "country" contain the majority of null values, we chose to treat eachmissing value is unavailable. Finally, we can see that there are no more missing values in the data frame.

Statistical Summary After Data Cleaning:

In [12]:
```
netflix_df.describe()
```

Out[12]:

	release_year
count	50.000000
mean	2014.340000
std	12.479551
min	1975.000000
25%	2014.000000
50%	2021.000000
75%	2021.000000
max	2021.000000

3. Non-Graphical Analysis

Non-Graphical Analysis involves calculating the summary statistics, without using pictorial or graphical representations. There are 3 main functions that Pandas library provide us, and I will be discussing about them. Those functions are:

1. info()
2. isna().sum() or isnull().sum()
3. describe()

In [14]:
```
# Checking the data using .head()
netflix_df.head()
```

Out[14]:

	show_id	type	title	director	cast	country	date_added	release_year	r
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No Cast	United States	September 25, 2021	2020	
1	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	
2	s11	TV Show	Vendetta: Truth, Lies and The Mafia	No Director	No Cast	Country Unavailable	September 24, 2021	2021	
3	s12	TV Show	Bangkok Breaking	Kongkiat Komesiri	Sukollawat Kanarot, Sushar Manaying, Pavarit M...	Country Unavailable	September 23, 2021	2021	
4	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	

1.info() mainly indicates the number of features, non-null count, and data type of each features. Additionally, it also shows the number of features in present in each data type(s). This helps us to determine how many numerical and categorical features we have.

In [15]:
```
netflix_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       50 non-null     object
 1   type          50 non-null     object
 2   title         50 non-null     object
 3   director      50 non-null     object
 4   cast          50 non-null     object
 5   country       50 non-null     object
 6   date_added    50 non-null     object
 7   release_year  50 non-null     int64 
 8   rating        50 non-null     object
 9   duration      50 non-null     object
 10  listed_in     50 non-null     object
 11  description   50 non-null     object
dtypes: int64(1), object(11)
memory usage: 4.8+ KB
```

2.Read The Description Of The Data

In [16]:
```
netflix_df.describe()
```

Out[16]:

	release_year
count	50.000000
mean	2014.340000
std	12.479551
min	1975.000000
25%	2014.000000
50%	2021.000000
75%	2021.000000
max	2021.000000

1. isna().sum() or isnull().sum()

In [18]:
```
netflix_df.T.apply(lambda x: x.isnull().sum(), axis = 1)
```

Out[18]:
```
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
```

In [19]:
```
netflix_movies_df = netflix_df[netflix_df.type == "Movie"]
netflix_shows_df = netflix_df[netflix_df.type == "TV Show"]
netflix_movies_df.head()
```

Out[19]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	No Cast	United States	September 25, 2021	2020
1	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021
4	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021
5	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Klara Castanho, Lucca Picon, Júlia Gomes, Marc...	Country Unavailable	September 22, 2021	2021
8	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in ...	Pedro de Echave García, Pablo Azorín Williams	No Cast	Country Unavailable	September 22, 2021	2020

In [20]:
```
netflix_shows_df.head()
```

Out[20]:

	show_id	type	title	director	cast	country	date_added	release_year	rati
2	s11	TV Show	Vendetta: Truth, Lies and The Mafia	No Director	No Cast	Country Unavailable	September 24, 2021	2021	T M
3	s12	TV Show	Bangkok Breaking	Kongkiat Komesiri	Sukollawat Kanarot, Sushar Manaying, Pavarit M...	Country Unavailable	September 23, 2021	2021	T M
6	s15	TV Show	Crime Stories: India Detectives	No Director	No Cast	Country Unavailable	September 22, 2021	2021	T M
7	s16	TV Show	Dear White People	No Director	Logan Browning, Brandon P. Bell, DeRon Horton,...	United States	September 22, 2021	2021	T M
9	s18	TV Show	Falsa identidad	No Director	Luis Ernesto Franco, Camila Sodi, Sergio Goyri...	Mexico	September 22, 2021	2020	T M

4: Exploratory Analysis and Visualization

Visual Analysis - Univariate, Bivariate after preprocessing of the data

Univariate analysis

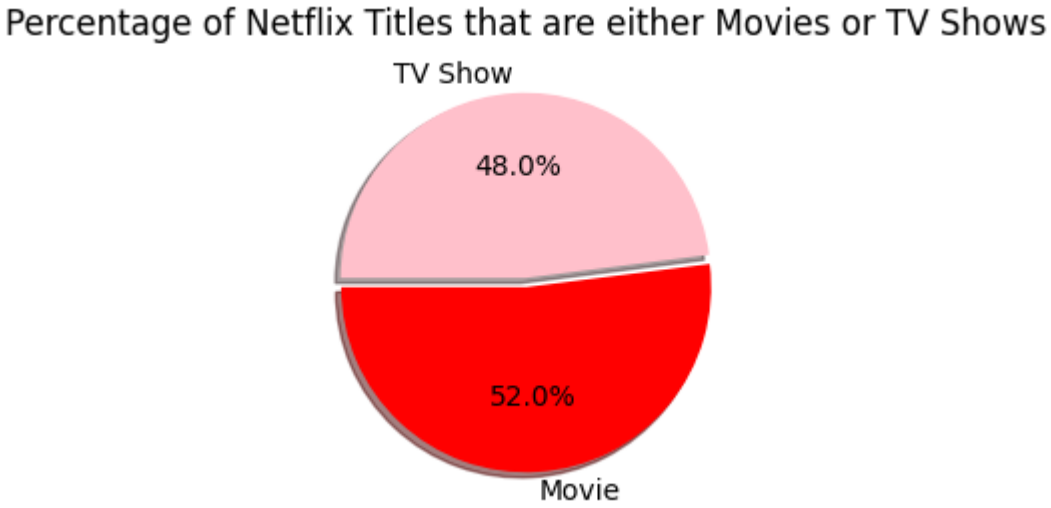
Analysis done based only on one variable. we are not going to the math behind these concepts, for now, let's see what these are in graphs.

A==>Pie plot:

Netflix Content By Type

Analysis entire Netflix dataset consisting of both movies and shows. Let’s compare the total number of movies and shows in this dataset to know which one is the majority

In [21]:
```
plt.figure(figsize=(6,3))
plt.title("Percentage of Netflix Titles that are either Movies or TV Shows")
g=plt.pie(netflix_df.type.value_counts(),explode=(0.025,0.025),
labels=netflix_df.type.value_counts().index, colors=['red','pink'],autopct='%1.1f%%',
startangle=180,shadow=True)
plt.show()
```



There are far more movie titles (52.0%) that TV shows titles (48.0%) in terms of title.

2. Amount of Content as a Function of Time:

Lineplot we will explore the amount of content Netflix has added throughout the previous years. Since we are interested in when Netflix added the title onto their platform, we will add a “year_added” column to show the date from the “date_added” columns.

In [22]:
```
# Handle potential inconsistencies in date formatting
netflix_df["year_added"] = pd.to_datetime(netflix_df.date_added.str.strip
(), format='%B %d, %Y', errors='coerce').dt.year
netflix_movies_df["year_added"] = pd.to_datetime(netflix_movies_df.date_add
ed.str.strip(), format='%B %d, %Y', errors='coerce').dt.year
netflix_shows_df["year_added"] = pd.to_datetime(netflix_shows_df.date_adde
d.str.strip(), format='%B %d, %Y', errors='coerce').dt.year
netflix_year_df = netflix_df.year_added.value_counts().to_frame().reset_ind
ex().rename(columns={"index": "count",
"year_added":"year"})
netflix_year_df = netflix_year_df[netflix_year_df['year'] != 2020]
print(netflix_year_df)
```

```
   year  count
0  2021     50

C:\Users\hp\AppData\Local\Temp\ipykernel_19828\3484994821.py:3: SettingWit
hCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  netflix_movies_df["year_added"] = pd.to_datetime(netflix_movies_df.date_
added.str.strip(), format='%B %d, %Y', errors='coerce').dt.year
C:\Users\hp\AppData\Local\Temp\ipykernel_19828\3484994821.py:4: SettingWit
hCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  netflix_shows_df["year_added"] = pd.to_datetime(netflix_shows_df.date_ad
ded.str.strip(), format='%B %d, %Y', errors='coerce').dt.year
```

In [23]:
```
movies_year_df =netflix_movies_df.year_added.value_counts().to_frame().rese
t_index().rename(columns={"index":
"count", "year_added":"year"})
movies_year_df = movies_year_df[movies_year_df['year'] != 2020]
movies_year_df
```

Out[23]:

	year	count
0	2021	26

In [24]:
```
shows_year_df = netflix_shows_df.year_added.value_counts().to_frame().reset
_index().rename(columns={"index":
"count", "year_added":"year"})
shows_year_df = shows_year_df[shows_year_df['year'] != 2020]
shows_year_df
```

Out[24]:

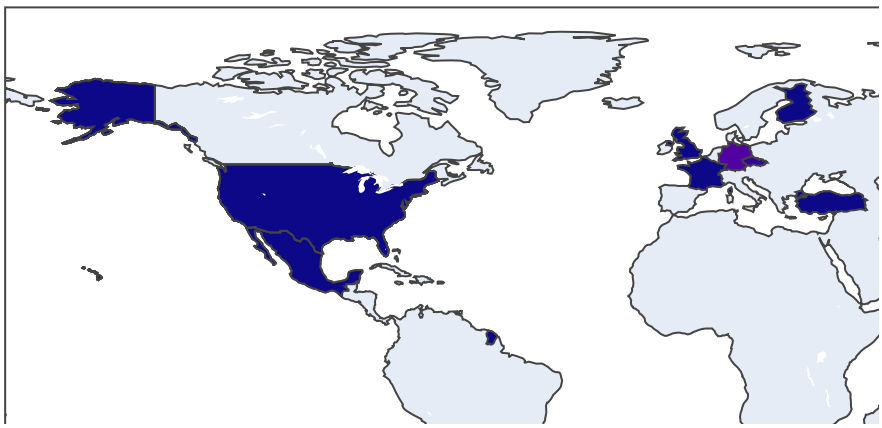
	year	count
0	2021	24

3. Exploring the countries contribution with the most content of Netflix.

Next is exploring the countries by the amount of the produces content of Netflix. We need to separate all countries within a film before analysing it, then removing titles with no countries available.

In [26]:
```
import plotly.graph_objects as go
from plotly.offline import init_notebook_mode, iplot
```

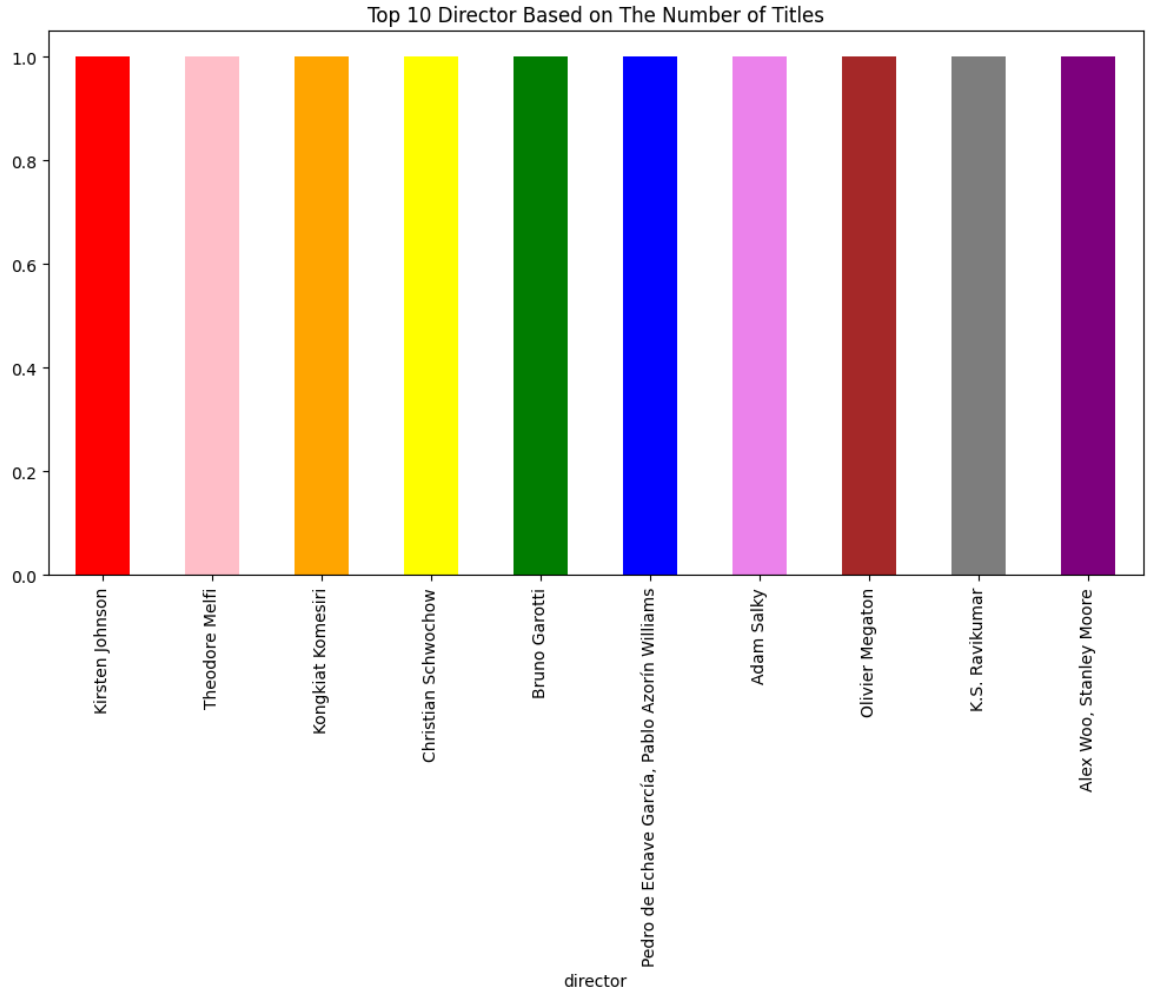
In [27]:
```
filtered_countries = netflix_df.set_index('title').country.str.split(', ',
expand=True).stack().reset_index(level=1, drop=True);
filtered_countries = filtered_countries[filtered_countries != 'Country Unav
ailable']
iplot([go.Choropleth(locationmode='country names',locations=filtered_countr
ies,z=filtered_countries.value_counts())])
```



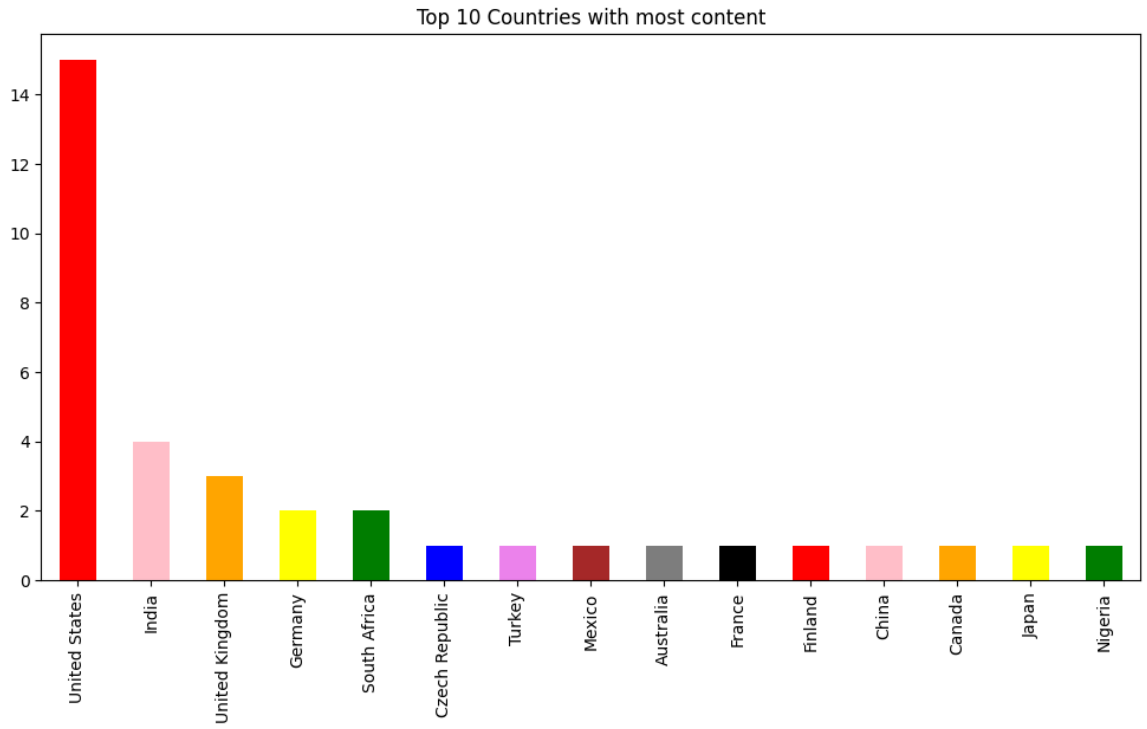
4. Top Directors on Netflix

To know the most popular director, we can visualize it.

In [28]:
```
plt.figure(figsize=(12,6))
netflix_df[netflix_df["director"] != "No Director"]["director"].value_counts().head(10).plot(kind="bar",color=['red','pink','orange','yellow','green','blue','violet','brown','gray','purple'])
plt.title("Top 10 Director Based on The Number of Titles")
plt.show()
```



In [29]:
```
plt.figure(figsize=(12,6))
filtered_countries.value_counts().head(15).plot(kind="bar",color=['red','pink','orange','yellow','green','blue','violet','brown','gray','black'])
plt.title("Top 10 Countries with most content")
plt.show()
```



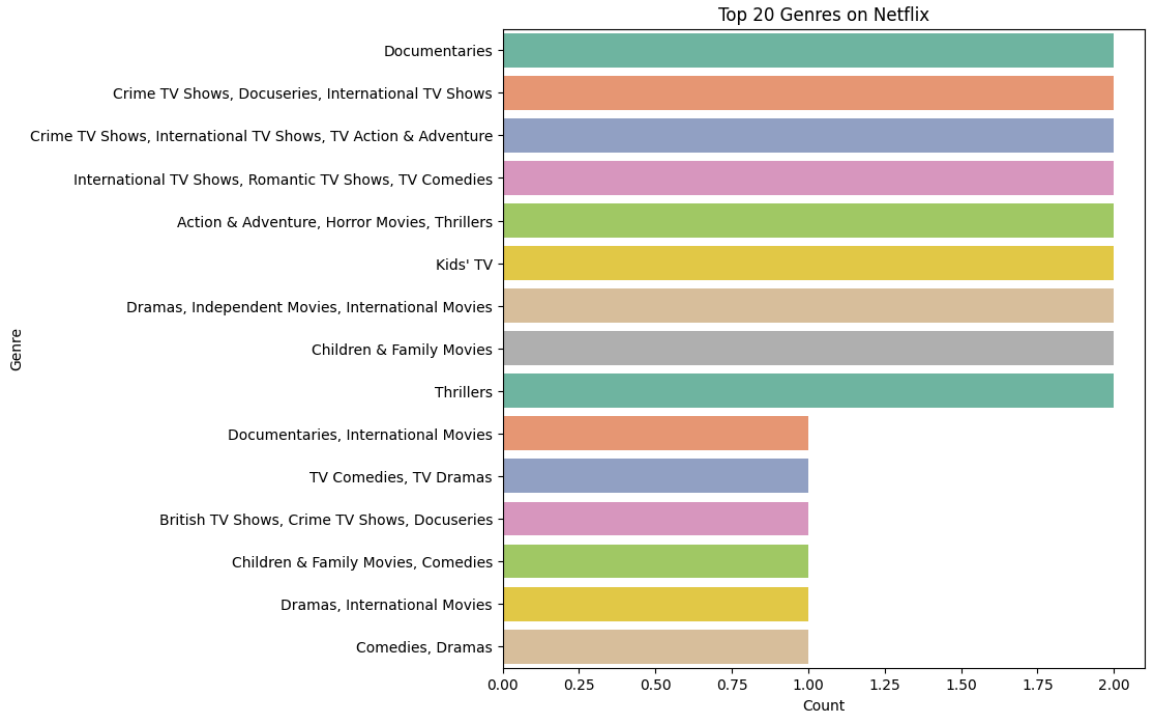
- The United States has the highest number of titles on Netflix, followed by India and the United Kingdom.
- This suggests that Netflix has a strong focus on content produced in these countries.
- However, it's important to note that this data may not reflect the popularity of content in different countries.
- For example, a title produced in the United States may be more popular in India than a title produced in India.

5. Top 20 Genres on Netflix: Count Plot

In [31]:
```
plt.figure(figsize=(8,8))
ax = sns.countplot(y="listed_in", data=netflix_df, order=netflix_df.listed_
in.value_counts().index[0:15], palette="Set2")
plt.title("Top 20 Genres on Netflix")
plt.xlabel("Count")
plt.ylabel("Genre")
plt.show()
```

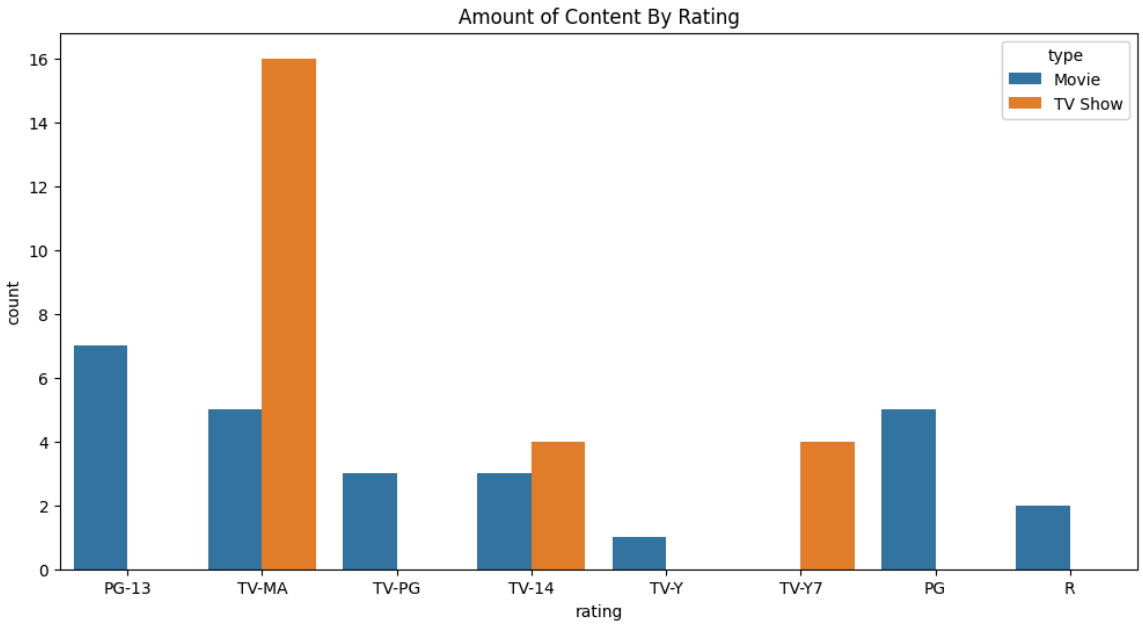
```
C:\Users\hp\AppData\Local\Temp\ipykernel_19828\2902980768.py:2: FutureWarn
ing:

Passing `palette` without assigning `hue` is deprecated and will be remove
d in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for
the same effect.
```



Amount of Content By Rating

In [32]:
```
plt.figure(figsize=(12,6))
sns.countplot(x='rating',hue='type',data=netflix_df)
plt.title("Amount of Content By Rating")
plt.show()
```

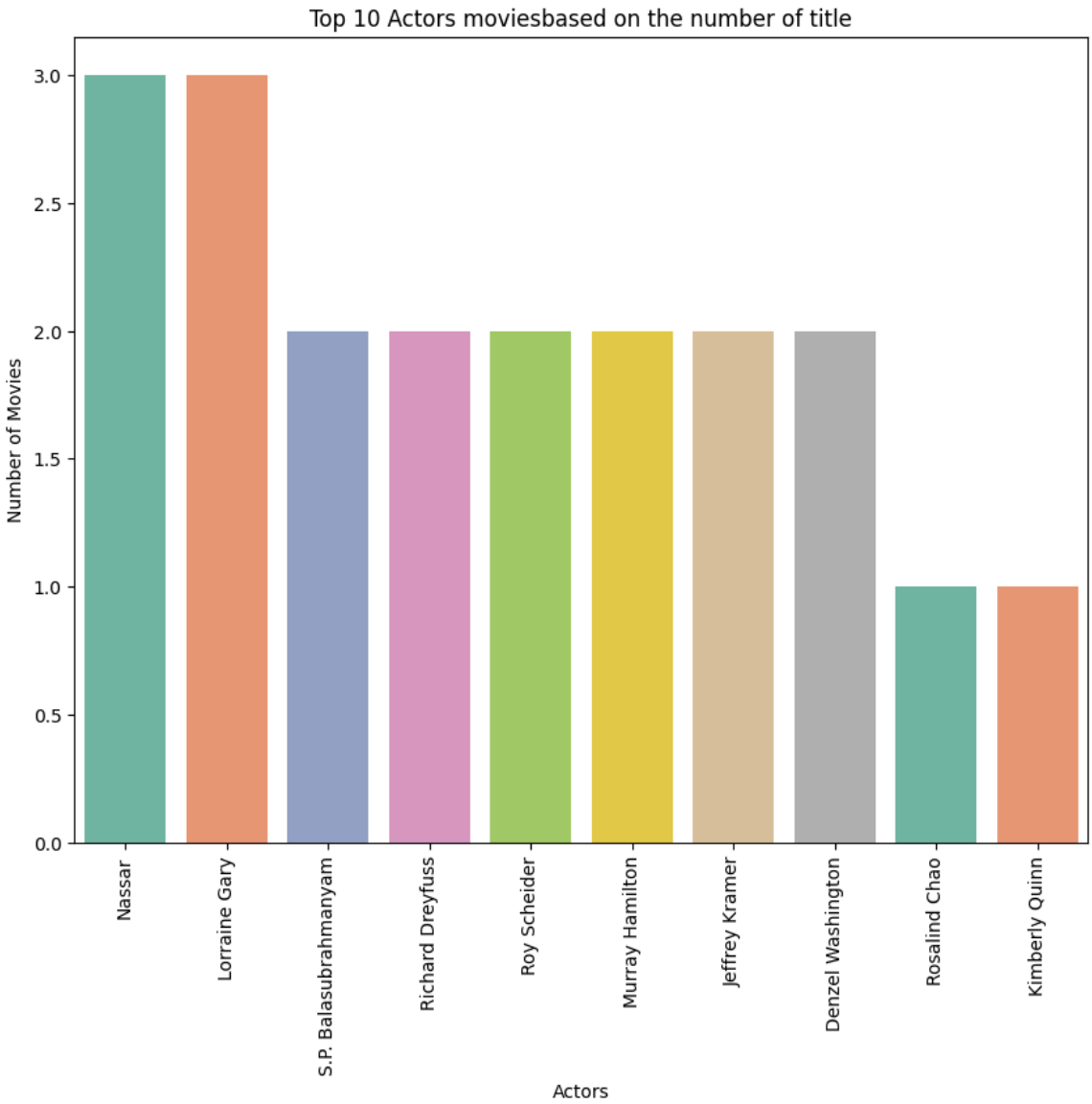


In [33]:
```
cast_shows = netflix_df[netflix_df["cast"] != "No Cast"].set_index("titl
e").cast.str.split(", ", expand=True).stack().reset_index(level=1, drop=Tru
e)
top_10_actors = cast_shows.value_counts().head(10)

plt.figure(figsize=(10, 8))
sns.barplot(x=top_10_actors.index, y=top_10_actors.values,  palette='Set2')
plt.xticks(rotation=90)
plt.xlabel("Actors")
plt.ylabel("Number of Movies")
plt.title("Top 10 Actors moviesbased on the number of title")
plt.show()
```

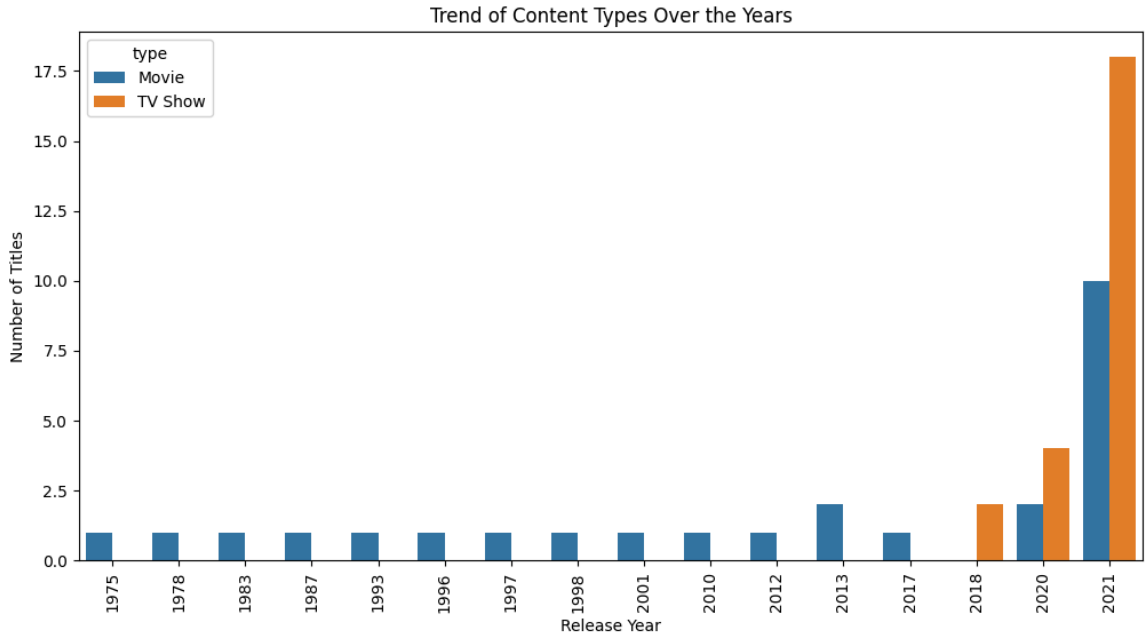
```
C:\Users\hp\AppData\Local\Temp\ipykernel_19828\1837082774.py:5: FutureWarn
ing:

Passing `palette` without assigning `hue` is deprecated and will be remove
d in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for
the same effect.
```



The top actor on Netflix Movies, based on the number of titles, is Nassar ,Lorraine Gary and S.P Balasubrahmanyam .

In [34]:
```
plt.figure(figsize=(12,6))
sns.countplot(x='release_year', hue='type', data=netflix_df)
plt.xticks(rotation=90)
plt.title('Trend of Content Types Over the Years')
plt.xlabel('Release Year')
plt.ylabel('Number of Titles')
plt.show()
```



In [ ]: