

Netflix Data Analysis

Breakdown of this notebook:

1. Importing Libraries

2. Loading the dataset

3. Data Cleaning:

- Deleting redundant columns.
- Dropping duplicates.
- Cleaning individual columns.
- Remove the NaN values from the dataset
- Some Transformations

4. Data Visualization: Using plots to find relations between the features.

- Type: Movie and TV Shows
- Rating
- Relation between Type and Rating

5. Word Cloud

- Country
- Cast
- Director
- Category

Importing Libraries

In [1]:

```
import numpy as np # linear algebra
import pandas as pd # data processing
import matplotlib.pyplot as plt
import seaborn as sns
# Input data files are available in the "../input/" directory.
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

/kaggle/input/netflix-shows/netflix_titles.csv

Loading the Dataset

In [2]:

```
netflix=pd.read_csv('../input/netflix-shows/netflix_titles.csv')
```

First 10 values

In [3]:

```
netflix.head(10)
```

Out[3]:

	show_id	type	title	director	cast	country	date_added	release_date
0	81145628	Movie	Norm of the North: King Sized Adventure	Richard Finn, Tim Maltby	Alan Marriott, Andrew Toth, Brian Dobson, Cole...	United States, India, South Korea, China	September 9, 2019	2019
1	80117401	Movie	Jandino: Whatever it Takes	NaN	Jandino Asporaat	United Kingdom	September 9, 2016	2016
2	70234439	TV Show	Transformers Prime	NaN	Peter Cullen, Sumalee Montano, Frank Welker, J...	United States	September 8, 2018	2013
3	80058654	TV Show	Transformers: Robots in Disguise	NaN	Will Friedle, Darren Criss, Constance Zimmer, ...	United States	September 8, 2018	2016
4	80125979	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	September 8, 2017	2017
5	80163890	TV Show	Apaches	NaN	Alberto Ammann, Eloy Azorín, Verónica Echegui,...	Spain	September 8, 2017	2016
6	70304989	Movie	Automata	Gabe Ibáñez	Antonio Banderas, Dylan McDermott, Melanie Gri...	Bulgaria, United States, Spain, Canada	September 8, 2017	2014
7	80164077	Movie	Fabrizio Copano: Solo pienso en mi	Rodrigo Toro, Francisco Schultz	Fabrizio Copano	Chile	September 8, 2017	2017
8	80117902	TV Show	Fire Chasers	NaN	NaN	United States	September 8, 2017	2017

	show_id	type	title	director	cast	country	date_added	relea
9	70304990	Movie	Good People	Henrik Ruben Genz	James Franco, Kate Hudson, Tom Wilkinson, Omar...	United States, United Kingdom, Denmark, Sweden	September 8, 2017	2014



Shape of the dataset

In [4]:

```
netflix.shape
```

Out[4]:

```
(6234, 12)
```

Print the name of columns

In [5]:

```
netflix.columns
```

Out[5]:

```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country',
      'date_added',
      'release_year', 'rating', 'duration', 'listed_in', 'descript
      ion'],
      dtype='object')
```

Check for NULL Values

In [6]:

```
netflix.isnull().sum()
```

Out[6]:

show_id	0
type	0
title	0
director	1969
cast	570
country	476
date_added	11
release_year	0
rating	10
duration	0
listed_in	0
description	0

dtype: int64

Check unique values

In [7]:

```
netflix.nunique()
```

Out[7]:

show_id	6234
type	2
title	6172
director	3301
cast	5469
country	554
date_added	1524
release_year	72
rating	14
duration	201
listed_in	461
description	6226

dtype: int64

Check for Duplicate values

In [8]:

```
netflix.duplicated().sum()
```

Out[8]:

0

No duplicate values present.

Make a copy of the dataset

In [9]:

```
df = netflix.copy()
```

In [10]:

```
df.shape
```

Out[10]:

(6234, 12)

Drop NULL values

In [11]:

```
df=df.dropna()  
df.shape
```

Out[11]:

(3774, 12)

Print first 10 values

In [12]:

```
df.head(10)
```

Out[12]:

	show_id	type	title	director	cast	country	date_added
0	81145628	Movie	Norm of the North: King Sized Adventure	Richard Finn, Tim Maltby	Alan Marriott, Andrew Toth, Brian Dobson, Cole...	United States, India, South Korea, China	September 9, 2019
4	80125979	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	September 8, 2017
6	70304989	Movie	Automata	Gabe Ibáñez	Antonio Banderas, Dylan McDermott, Melanie Gri...	Bulgaria, United States, Spain, Canada	September 8, 2017
7	80164077	Movie	Fabrizio Copano: Solo pienso en mi	Rodrigo Toro, Francisco Schultz	Fabrizio Copano	Chile	September 8, 2017
9	70304990	Movie	Good People	Henrik Ruben Genz	James Franco, Kate Hudson, Tom Wilkinson, Omar...	United States, United Kingdom, Denmark, Sweden	September 8, 2017
11	70299204	Movie	Kidnapping Mr. Heineken	Daniel Alfredson	Jim Sturgess, Sam Worthington, Ryan Kwanten, A...	Netherlands, Belgium, United Kingdom, United S...	September 8, 2017
19	80057969	Movie	Love	Gaspar Noé	Karl Glusman, Klara Kristin, Aomi Muyock, Ugo ...	France, Belgium	September 8, 2017
20	80060297	Movie	Manhattan Romance	Tom O'Brien	Tom O'Brien, Katherine Waterston, Caitlin Fitz...	United States	September 8, 2017
21	80046728	Movie	Moonwalkers	Antoine Bardou-Jacquet	Ron Perlman, Rupert Grint, Robert Sheehan, Ste...	France, Belgium	September 8, 2017

	show_id	type	title	director	cast	country	date_added
23	70304988	Movie	Stonehearst Asylum	Brad Anderson	Kate Beckinsale, Jim Sturgess, David Thewlis, ...	United States	September 8, 2017



Convert Date Time format

In [13]:

```
df["date_added"] = pd.to_datetime(df['date_added'])
df['day_added'] = df['date_added'].dt.day
df['year_added'] = df['date_added'].dt.year
df['month_added'] = df['date_added'].dt.month
df['year_added'].astype(int);
df['day_added'].astype(int);
```

In [14]:

```
df.head(10)
```

Out[14]:

	show_id	type	title	director	cast	country	date_added
0	81145628	Movie	Norm of the North: King Sized Adventure	Richard Finn, Tim Maltby	Alan Marriott, Andrew Toth, Brian Dobson, Cole...	United States, India, South Korea, China	2019-09-09
4	80125979	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	2017-09-08
6	70304989	Movie	Automata	Gabe Ibáñez	Antonio Banderas, Dylan McDermott, Melanie Gri...	Bulgaria, United States, Spain, Canada	2017-09-08
7	80164077	Movie	Fabrizio Copano: Solo pienso en mi	Rodrigo Toro, Francisco Schultz	Fabrizio Copano	Chile	2017-09-08
9	70304990	Movie	Good People	Henrik Ruben Genz	James Franco, Kate Hudson, Tom Wilkinson, Omar...	United States, United Kingdom, Denmark, Sweden	2017-09-08
11	70299204	Movie	Kidnapping Mr. Heineken	Daniel Alfredson	Jim Sturgess, Sam Worthington, Ryan Kwanten, A...	Netherlands, Belgium, United Kingdom, United S...	2017-09-08
19	80057969	Movie	Love	Gaspar Noé	Karl Glusman, Klara Kristin, Aomi Muyock, Ugo ...	France, Belgium	2017-09-08
20	80060297	Movie	Manhattan Romance	Tom O'Brien	Tom O'Brien, Katherine Waterston, Caitlin Fitz...	United States	2017-09-08
21	80046728	Movie	Moonwalkers	Antoine Bardou-Jacquet	Ron Perlman, Rupert Grint, Robert Sheehan, Ste...	France, Belgium	2017-09-08

	show_id	type	title	director	cast	country	date_added	
23	70304988	Movie	Stonehearst Asylum	Brad Anderson	Kate Beckinsale, Jim Sturgess, David Thewlis, ...	United States	2017-09-08	

Data Visualization

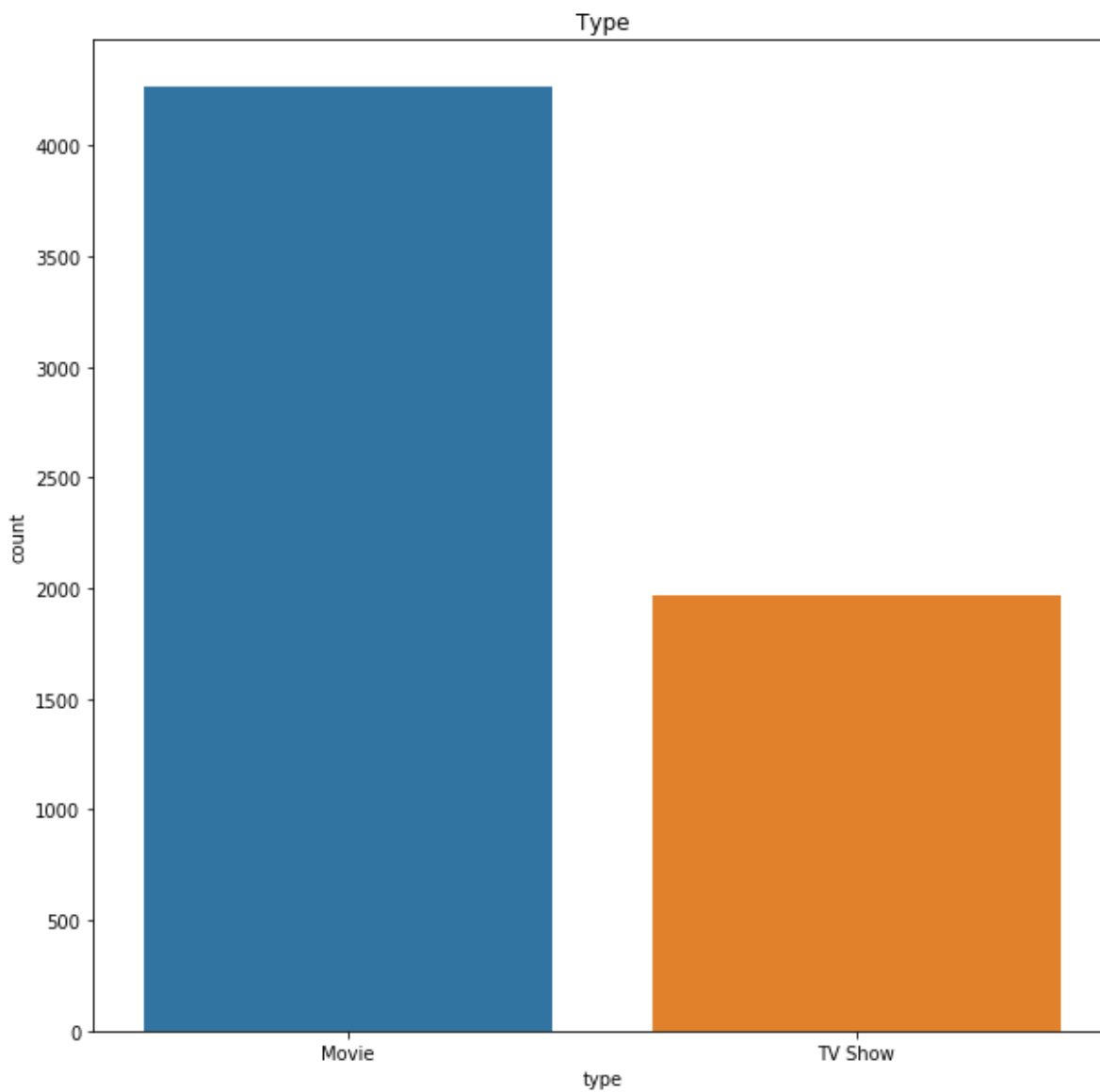
Type: Movie and TV Shows

In [15]:

```
sns.countplot(netflix['type'])  
fig = plt.gcf()  
fig.set_size_inches(10,10)  
plt.title('Type')
```

Out[15]:

Text(0.5, 1.0, 'Type')



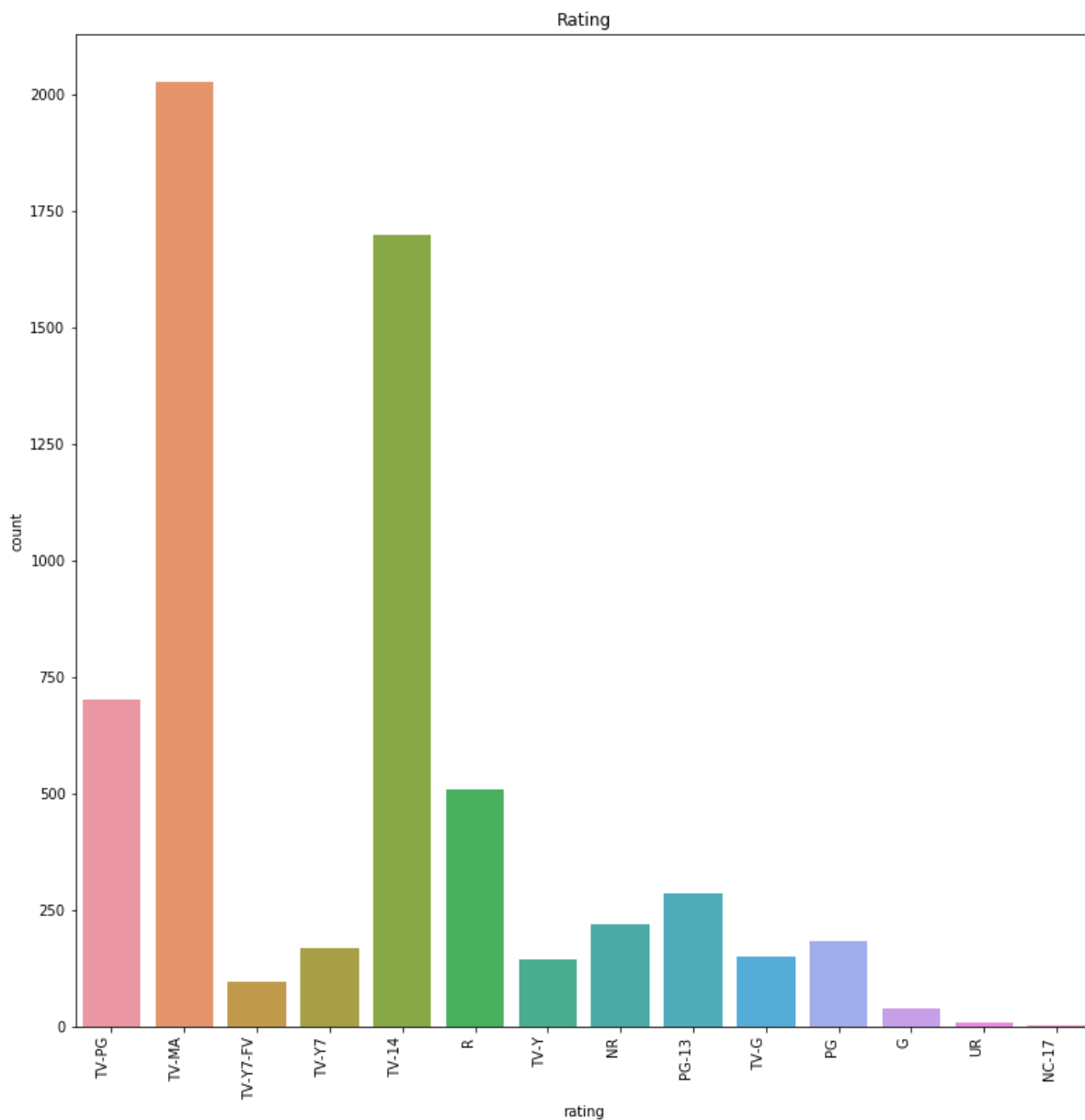
Rating of shows and movies

In [16]:

```
sns.countplot(netflix['rating'])
sns.countplot(netflix['rating']).set_xticklabels(sns.countplot(netflix['rating']).get_xticklabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Rating')
```

Out[16]:

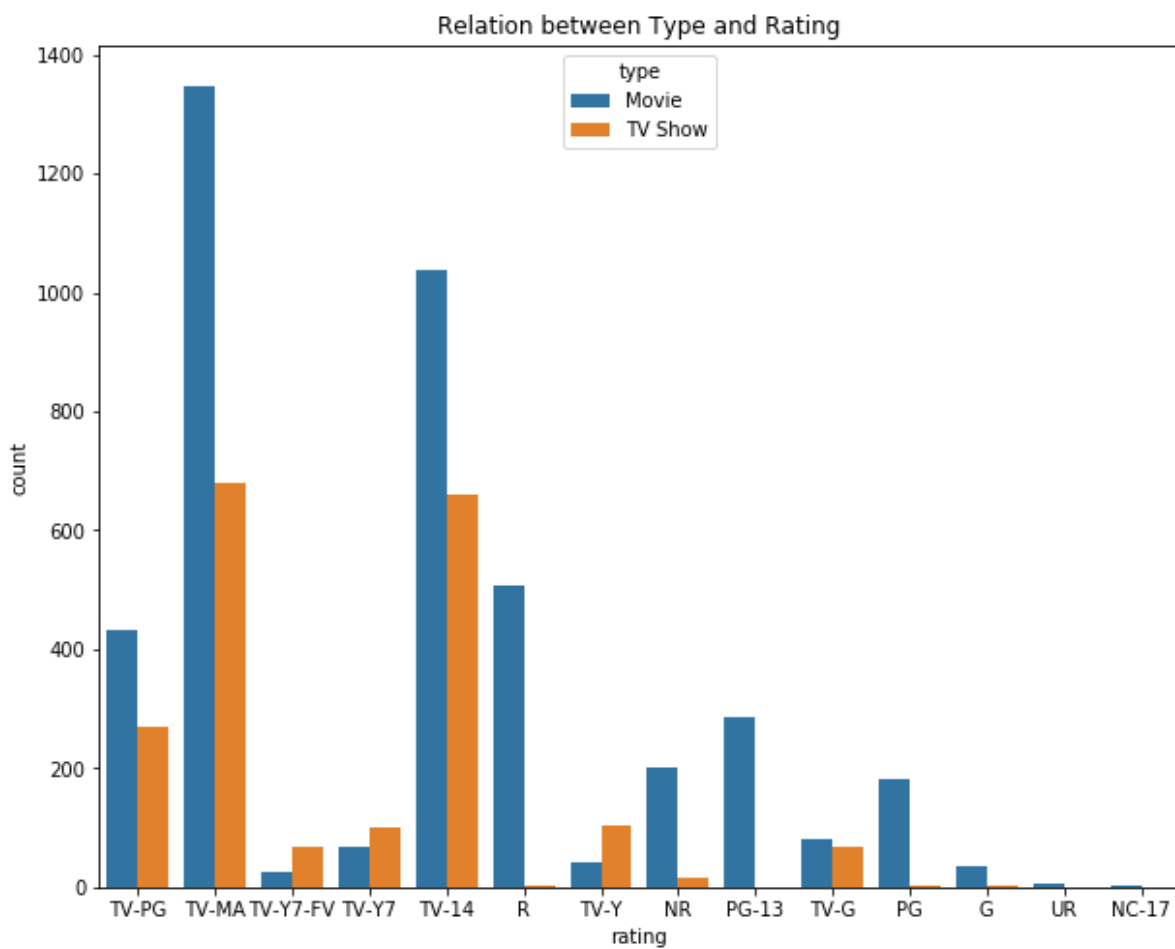
Text(0.5, 1.0, 'Rating')



Relation between Type and Rating

In [17]:

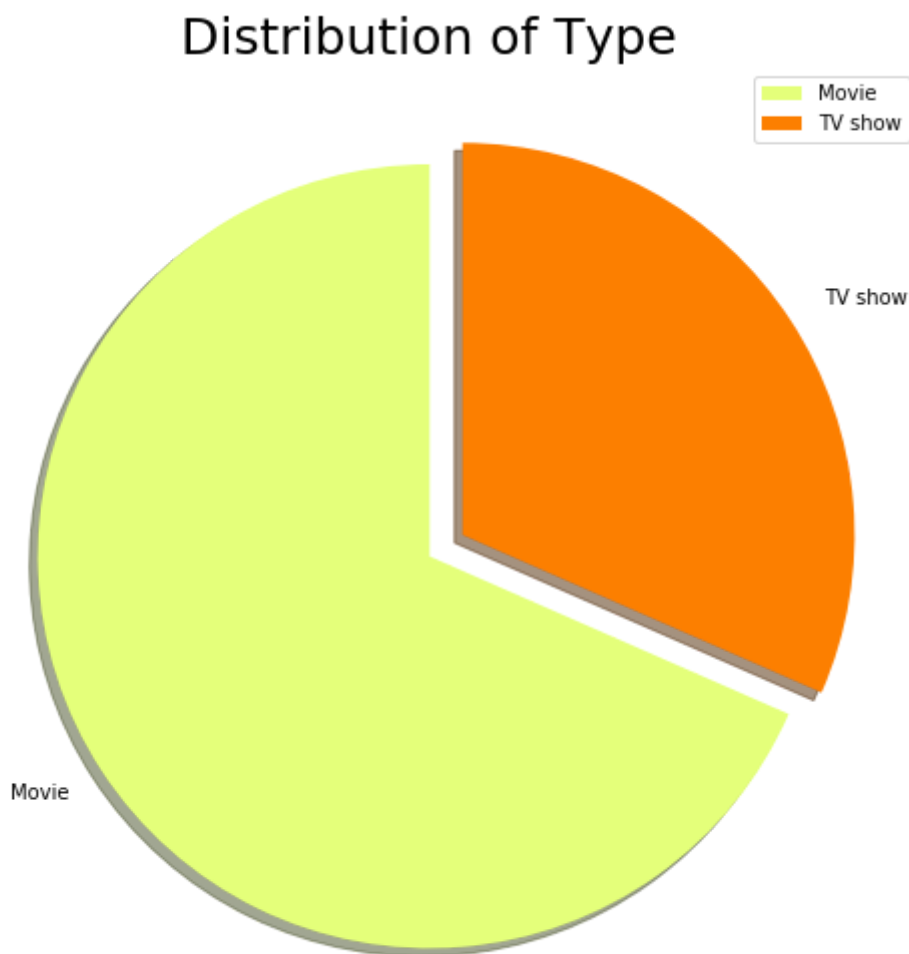
```
plt.figure(figsize=(10,8))
sns.countplot(x='rating',hue='type',data=netflix)
plt.title('Relation between Type and Rating')
plt.show()
```



Pie-chart for the Type: Movie and TV Shows

In [18]:

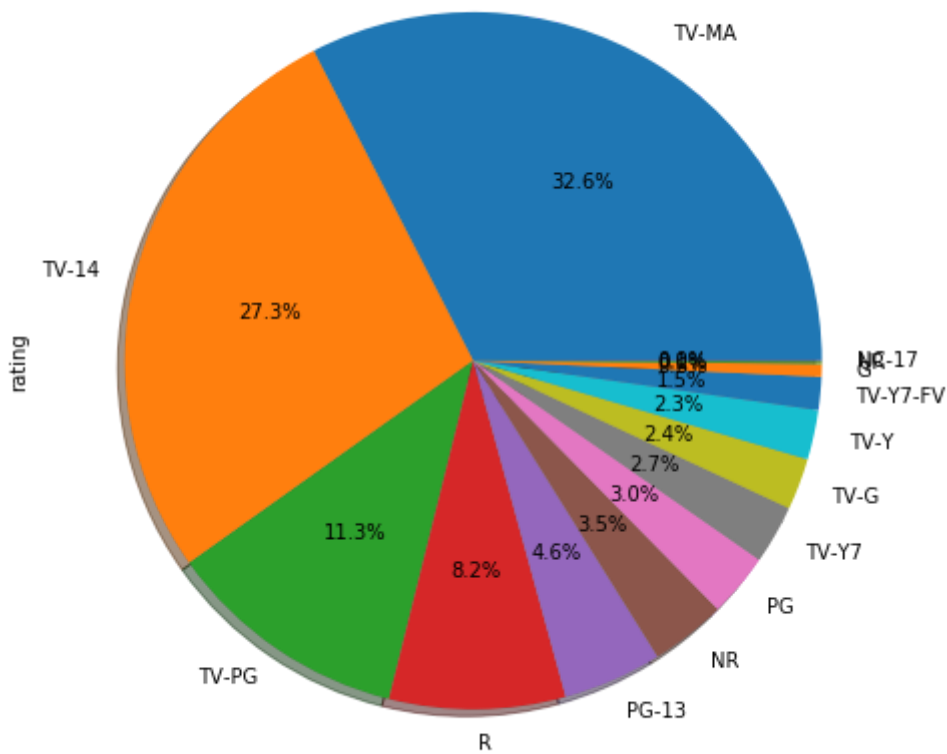
```
labels = ['Movie', 'TV show']
size = netflix['type'].value_counts()
colors = plt.cm.Wistia(np.linspace(0, 1, 2))
explode = [0, 0.1]
plt.rcParams['figure.figsize'] = (9, 9)
plt.pie(size, labels=labels, colors = colors, explode = explode, shadow = True, s
tartangle = 90)
plt.title('Distribution of Type', fontsize = 25)
plt.legend()
plt.show()
```



Pie-chart for Rating

In [19]:

```
netflix['rating'].value_counts().plot.pie(autopct='%1.1f%%', shadow=True, figsize=(10,8))  
plt.show()
```



WordCloud

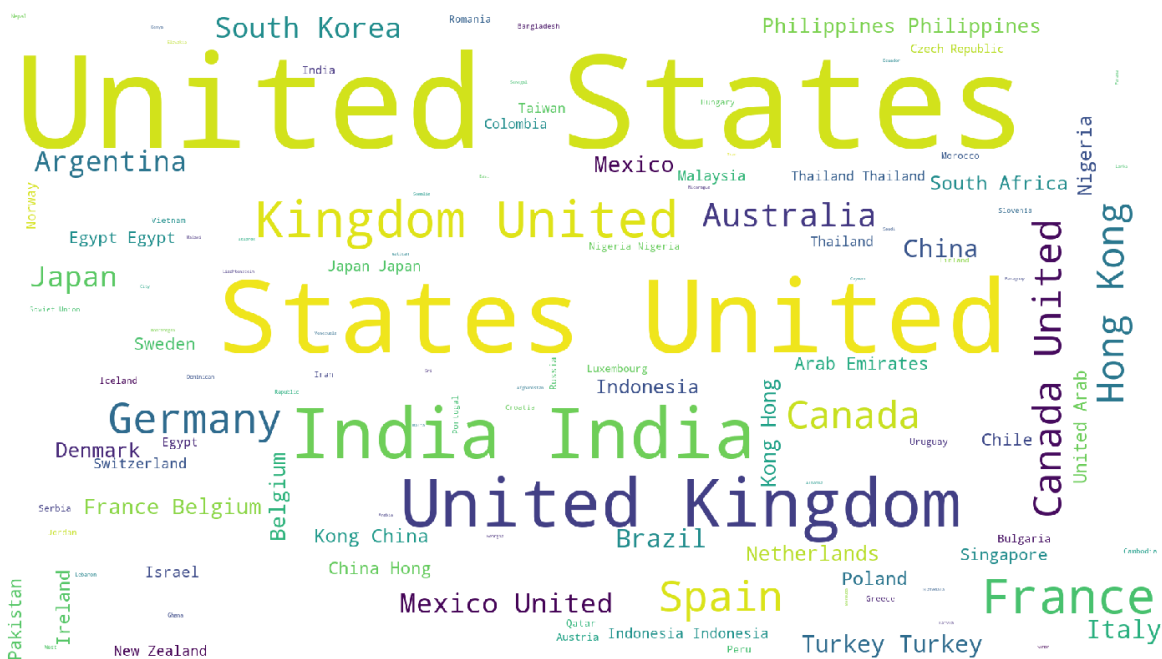
In [20]:

```
from wordcloud import WordCloud
```

Country

In [21]:

```
plt.subplots(figsize=(25,15))
wordcloud = WordCloud(
    background_color='white',
    width=1920,
    height=1080
).generate(" ".join(df.country))
plt.imshow(wordcloud)
plt.axis('off')
plt.savefig('country.png')
plt.show()
```

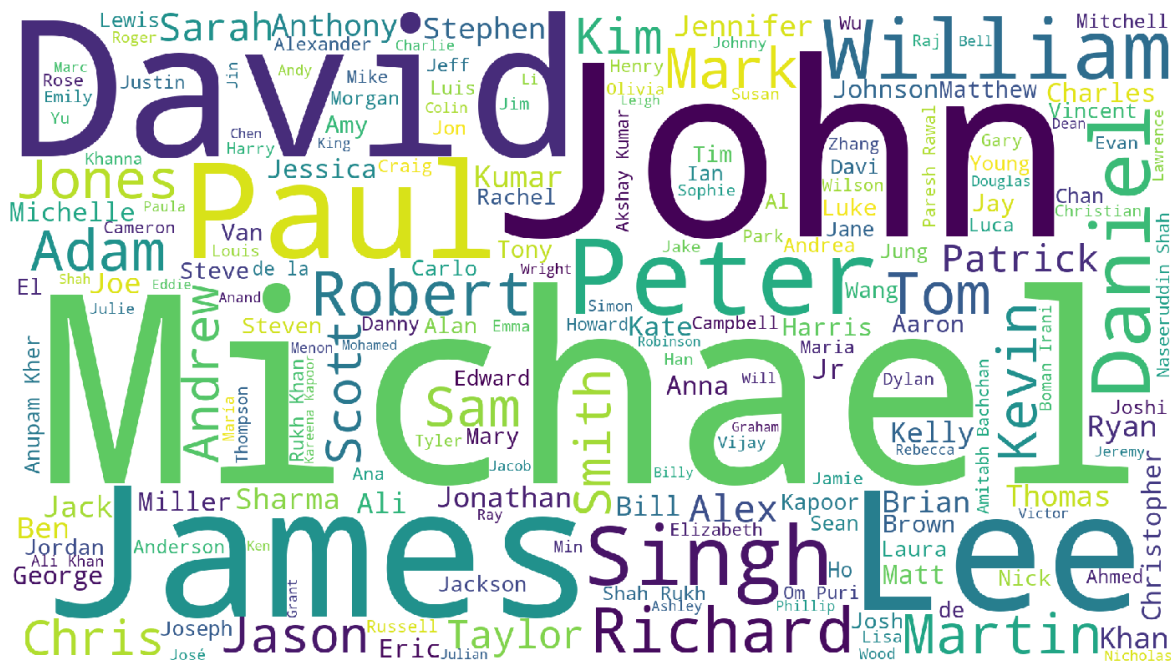


Cast in the Shows

In [22]:

```
plt.subplots(figsize=(25,15))
wordcloud = WordCloud(
    background_color='white',
    width=1920,
    height=1080
).generate(" ".join(df.cast))

plt.imshow(wordcloud)
plt.axis('off')
plt.savefig('cast.png')
plt.show()
```

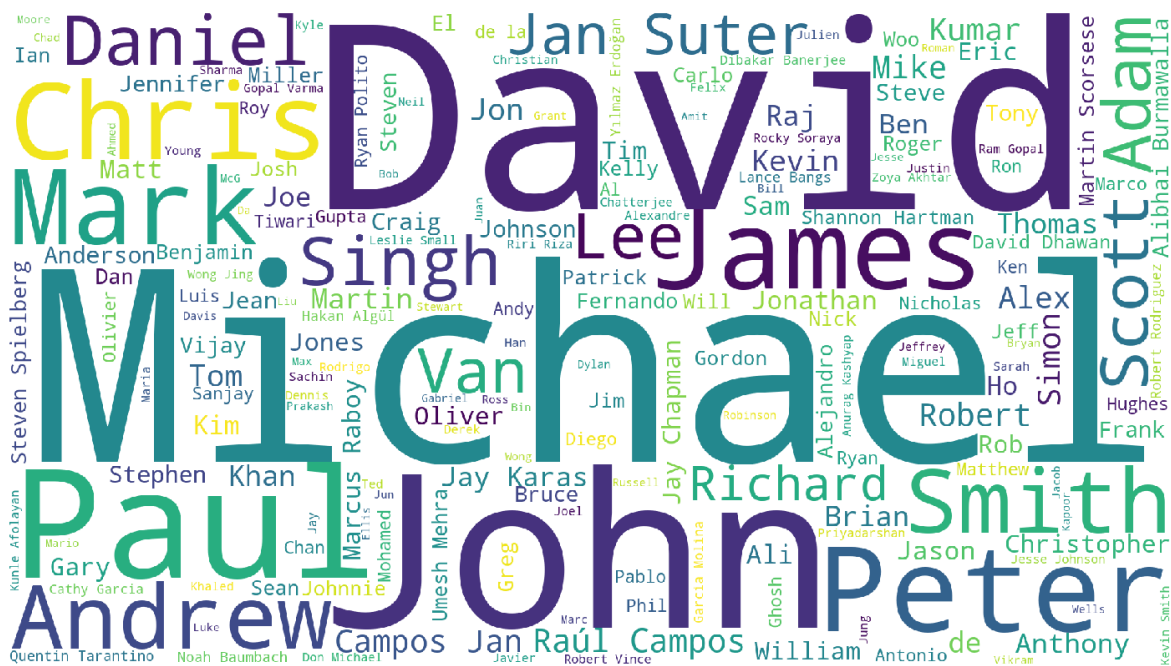


Directors

In [23]:

```
plt.subplots(figsize=(25,15))
wordcloud = WordCloud(
    background_color='white',
    width=1920,
    height=1080
).generate(" ".join(df.director))

plt.imshow(wordcloud)
plt.axis('off')
plt.savefig('director.png')
plt.show()
```



Categories

In [24]:

```
plt.subplots(figsize=(25,15))
wordcloud = WordCloud(
    background_color='white',
    width=1920,
    height=1080
).generate(" ".join(df.listed_in))
plt.imshow(wordcloud)
plt.axis('off')
plt.savefig('category.png')
plt.show()
```

