

Netflix Titles Analysis

```
#importing the pandas library to carry out the analysis
import pandas as pd
```

```
#creating a data frame into which the data is being read
df = pd.read_csv("netflix_titles.csv")
```

```
#displaying the contents of the dataframe
df.head(5)
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons
1	s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min

```
#shape of the dataframe that indicates the number of rows and columns
df.shape
```

(7787, 12)

```
#size of the dataframe that indicates the total number of elements in it given by rows x columns
df.size
```

93444

```
#provides information regarding the number of dimensions in the dataframe
df.ndim
```

2

```
#display all the columns of the dataframe
df.columns
```

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

```
#the same columns in the form of a list can be given by
list(df.columns)
```

```
['show_id',
 'type',
 'title',
 'director',
 'cast',
 'country',
```

```
'date_added',
'release_year',
'rating',
'duration',
'listed_in',
'description']
```

```
#summary statistics of the data
df.describe()
```

	release_year
count	7787.000000
mean	2013.932580
std	8.757395
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2018.000000
max	2021.000000

```
#correlation
df.corr()
```

	release_year
release_year	1.0

```
#what unique values are present in a given column
df["type"].unique()

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

```
#filtering data based on the needs (single query)
df[df["country"] == 'West Germany']
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
2757	s2758	Movie	Hitler - A Career	Joachim Fest, Christian Herrendoerfer	NaN	West Germany	February 10, 2017	1977	TV-MA	150 min

```
#filtering data based on the needs (combining multiple queries)
df[(df["country"] == "Singapore") & (df["rating"] == "TV-MA" )]
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	du
2093	s2094	Movie	Fakkah Fuzz: Almost Banned	Michael McKay	Fakkah Fuzz	Singapore	January 26, 2018	2018	TV-MA	
2956	s2957	Movie	Imperfect	Steve Cheng	Edwin Goh, Ian Fang Weijie, Phua	Singapore	December 20, 2018	2012	TV-MA	

```
#getting all the date values in a given row
computer_date = pd.to_datetime(df["date_added"])
```

```
#displaying all the dates of the date_added column
computer_date
```

```
0      2020-08-14
1      2016-12-23
2      2018-12-20
3      2017-11-16
4      2020-01-01
...
7782   2020-10-19
7783   2019-03-02
7784   2020-09-25
7785   2020-10-31
7786   2020-03-01
Name: date_added, Length: 7787, dtype: datetime64[ns]
```

```
TV      The Truth      Lim.
#importing the date functionality from datetime module
from datetime import date
```

```
#creating a variable to store the date and displaying
saturday_date = date.today()
saturday_date
```

```
datetime.date(2023, 2, 18)
```

```
#total number of null values in a column of the dataframe
df["date_added"].isna().sum()
```

```
10
```

```
from datetime import datetime as dt
```

```
#replacing all the missing values in the column
df["date_added"].fillna(dt(2020, 1, 1), inplace = True)
```

```
#verifying that there are no missing values in the column
df["date_added"].isna().sum()
```

```
0
```

```
#splitting the column values as per the requirement
df[["part_1", "part_2"]] = df["date_added"].str.split(" ", expand = True)
```

```
df
```

NaN	Diana Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons	TV Shows, TV Dramas, TV Sci-Fi &...	where the inh...
Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min	Dramas, International Movies	devastating earthquake Mexico
Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min	Horror Movies, International Movies	When a recruit is de...
Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min	Action & Adventure, Independent Movies, Sci-Fi...	postapocalyptic world, robots
Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	123 min	Dramas	A brilliant of st... become
...
Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	2005	TV-MA	99 min	Dramas, International Movies	Lebanon War de Zozo o
Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrap pc worms h into
NaN	Nasty C	NaN	September 25, 2020	2019	TV-MA	44 min	Documentaries, International Movies, Music & M...	docum South rapper
NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	2019	TV-PG	1 Season	International TV Shows, Reality TV	Dessert Adriano looks
Sam Dunn	NaN	United Kingdom, Canada, United States	March 1, 2020	2019	TV-MA	90 min	Documentaries, Music & Musicals	docum delves i m

```
#renaming columns
df.rename(columns = {"part_2":"year"}, inplace = True)

df.head(5)
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons
1	s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min

```
list(df.columns)

['show_id',
 'type',
 'title',
 'director',
 'cast',
 'country',
 'date_added',
 'release_year',
 'rating',
 'duration',
 'listed_in',
 'description',
 'Day and month',
 'Year',
 'part_1',
 'year']

df["cast"].isna().sum()

718

df["cast"].fillna(value = "no cast", inplace = True)

df
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA
1	s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13
...
7782	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergl, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	2005	TV-MA
7783	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14

```
df.loc[7786]

show_id          s7787
type            Movie
title           ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS
director        Sam Dunn
cast            no cast
country          United Kingdom, Canada, United States
date_added      March 1, 2020
release_year    2019
rating          TV-MA
duration        90 min
listed_in       Documentaries, Music & Musicals
description     This documentary delves into the mystique behi...
Day and month   March 1
Year            2020
part_1          March 1
year            2020
Name: 7786, dtype: object
```

```
#lambda function
the_adder = lambda x: x + 10
```

```
the_adder(5)
```

15

```
df["genre_count"] = df["listed_in"].map(lambda x: len(x.split(",")))
```

```
df.head(5)
```

director	cast	country	date_added	release_year	rating	duration	listed_in	description	Date
NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...	Aug
Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min	Dramas, International Movies	After a devastating earthquake hits Mexico Cit...	Dec
Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min	Horror Movies, International Movies	When an army recruit is found dead, his fellow...	Dec
Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min	Action & Adventure, Independent Movies, Sci-Fi...	In a postapocalyptic world, rag-doll robots hi...	Nov
Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	123 min	Dramas	A brilliant group of students become card-coun...	Jan

```
movies_by_country = df.groupby("country")["show_id"].count().reset_index().sort_values(by = "show_id", ascending = False)
```

```
movies_by_country.loc[546]
```

```
country    United Kingdom, United States, Morocco
show_id    1
Name: 546, dtype: object
```

```
#concatenate is the equivalent of union in sql
```

```
second_dataset = pd.read_csv("netflix_titles_second.csv")
```

```
second_dataset.head(5)
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	14-Aug-20	2020	TV-MA	4 Seasons
1	s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar	Mexico	23-Dec-16	2016	TV-MA	93 min

```
new_dataset = pd.concat([df, second_dataset])

#we can see that the number has increased from before
new_dataset.shape

(8141, 17)

#merge method
netflix_merged = pd.merge(left = new_dataset , right = movies_by_country, how = "inner", left_on = ["country"], right_on = ["country"])

netflix_merged.head(3)
```

show_id_x	type	title	director	cast	country	date_added	release_year	
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020
1	s209	Movie	A new Capitalism	NaN	no cast	Brazil	June 12, 2018	2017
2	s219	TV Show	A Queen Is Born	Carla Barros	Gloria Groove, Alexia Twister	Brazil	November 11, 2020	2020

