

# Amazon Prime Videos Analysis Using Python

The Amazon Prime Videos set has the information about the Tv Shows & Movies.

The Data set available from Flexible which is a Third Party Amazon Prime Videos which engine , and available on Kaggle dataset for free.

## Import Library

```
In [1]: import pandas as pd
```

```
In [2]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns
```

```
C:\Users\Syed Arif\anaconda3\lib\site-packages\scipy\__init__.py:146: UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy (detected version 1.25.1)
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}")
```

## Uploading Csv file

```
In [3]: df = pd.read_csv(r"C:\Users\Syed Arif\Downloads\amazon_prime_titles.csv")
```

## Data Preprocessing

### .head()

head is used show to the By default = 5 rows in the dataset

In [4]: `df.head()`

Out[4]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	d
0	s1	Movie	The Grand Seduction	Don McKellar	Brendan Gleeson, Taylor Kitsch, Gordon Pinsent	Canada	March 30, 2021	2014	NaN	
1	s2	Movie	Take Care Good Night	Girish Joshi	Mahesh Manjrekar, Abhay Mahajan, Sachin Khedekar	India	March 30, 2021	2018	13+	
2	s3	Movie	Secrets of Deception	Josh Webber	Tom Sizemore, Lorenzo Lamas, Robert LaSardo, R...	United States	March 30, 2021	2017	NaN	
3	s4	Movie	Pink: Staying True	Sonia Anderson	Interviews with: Pink, Adele, Beyoncé, Britney...	United States	March 30, 2021	2014	NaN	
4	s5	Movie	Monster Maker	Giles Foster	Harry Dean Stanton, Kieran O'Brien, George Cos...	United Kingdom	March 30, 2021	1989	NaN	



## .tail()

tail is used to show rows by Descending order

In [5]: `df.tail()`

Out[5]:

	show_id	type	title	director	cast	country	date_added	release_year	rating
9663	s9664	Movie	Pride Of The Bowery	Joseph H. Lewis	Leo Gorcey, Bobby Jordan	NaN	NaN	1940	7
9664	s9665	TV Show	Planet Patrol	NaN	DICK VOSBURGH, RONNIE STEVENS, LIBBY MORRIS, M...	NaN	NaN	2018	13
9665	s9666	Movie	Outpost	Steve Barker	Ray Stevenson, Julian Wadham, Richard Brake, M...	NaN	NaN	2008	
9666	s9667	TV Show	Maradona: Blessed Dream	NaN	Esteban Recagno, Ezequiel Stremiz, Luciano Vit...	NaN	NaN	2021	TV M
9667	s9668	Movie	Harry Brown	Daniel Barber	Michael Caine, Emily Mortimer, Joseph Gilgun, ...	NaN	NaN	2010	

## .shape

It show the total no of rows & Column in the dataset

In [6]: `df.shape`

Out[6]: (9668, 12)

## .Columns

It show the no of each Column

```
In [7]: df.columns
```

```
Out[7]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',  
              'release_year', 'rating', 'duration', 'listed_in', 'description'],  
             dtype='object')
```

## .dtypes

This Attribute show the data type of each column

```
In [8]: df.dtypes
```

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

## .unique()

In a column, It show the unique value of specific column.

```
In [9]: df["director"].unique()
```

```
Out[9]: array(['Don McKellar', 'Girish Joshi', 'Josh Webber', ...,  
              'John-Paul Davidson', 'Stephen Warbeck', 'Emily Skye',  
              'Steve Barker'], dtype=object)
```

## .nunique()

It will show the total no of unique value from whole data frame

```
In [10]: df.nunique()
```

```
Out[10]: show_id      9668  
         type         2  
         title      9668  
         director   5774  
         cast       7927  
         country     86  
         date_added  84  
         release_year 100  
         rating      24  
         duration   219  
         listed_in   518  
         description 9414  
         dtype: int64
```

## .describe()

It show the Count, mean , median etc

```
In [11]: df.describe()
```

```
Out[11]:
```

	release_year
count	9668.000000
mean	2008.341849
std	18.922482
min	1920.000000
25%	2007.000000
50%	2016.000000
75%	2019.000000
max	2021.000000

## .value\_counts

It Shows all the unique values with their count

```
In [12]: df["director"].value_counts()
```

```
Out[12]: Mark Knight          113
Cannis Holder              61
Moonbug Entertainment      37
Jay Chapman                34
Arthur van Merwijk         30
...
Andrew Traucki             1
Karyn Kusama               1
K. Subash                  1
Robert Cuffley             1
Steve Barker               1
Name: director, Length: 5774, dtype: int64
```

## .isnull()

It shows the how many null values

```
In [13]: df.isnull()
```

```
Out[13]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
0	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False	True	False
4	False	False	False	False	False	False	False	False	True	False
...	...	...	...	...	...	...	...	...	...	...
9663	False	False	False	False	False	True	True	False	False	False
9664	False	False	False	True	False	True	True	False	False	False
9665	False	False	False	False	False	True	True	False	False	False
9666	False	False	False	True	False	True	True	False	False	False
9667	False	False	False	False	False	True	True	False	False	False

9668 rows × 12 columns

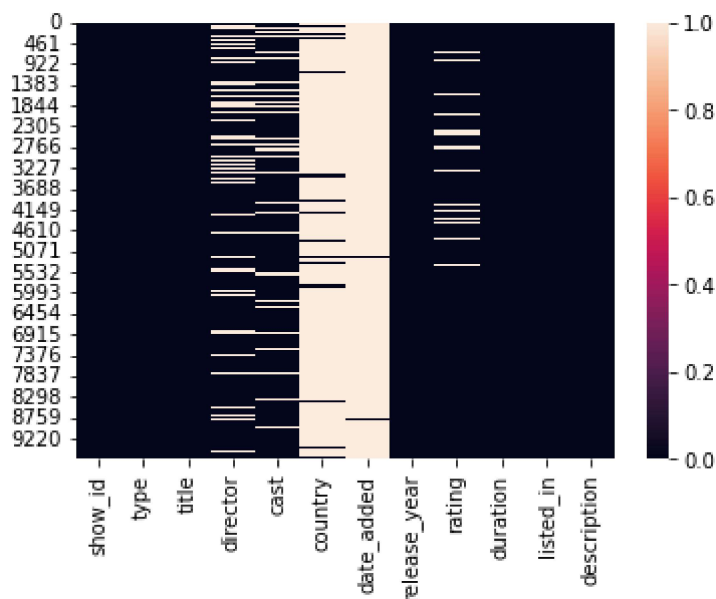


In [14]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9668 entries, 0 to 9667
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         9668 non-null   object
1   type            9668 non-null   object
2   title           9668 non-null   object
3   director        7586 non-null   object
4   cast            8435 non-null   object
5   country         672 non-null    object
6   date_added      155 non-null    object
7   release_year    9668 non-null   int64
8   rating          9331 non-null   object
9   duration        9668 non-null   object
10  listed_in       9668 non-null   object
11  description     9668 non-null   object
dtypes: int64(1), object(11)
memory usage: 906.5+ KB
```

**How many Null value present show all the null values in Heatmap.**

In [15]: sns.heatmap(df.isnull())  
plt.show()



**In which year Heighest number of Tv shows and Movies were release.**

```
In [16]: df['release_year'].value_counts()
```

```
Out[16]: 2021    1442
          2020     962
          2019     929
          2018     623
          2017     562
          ...
          1922      2
          1926      2
          1924      1
          1923      1
          1927      1
          Name: release_year, Length: 100, dtype: int64
```

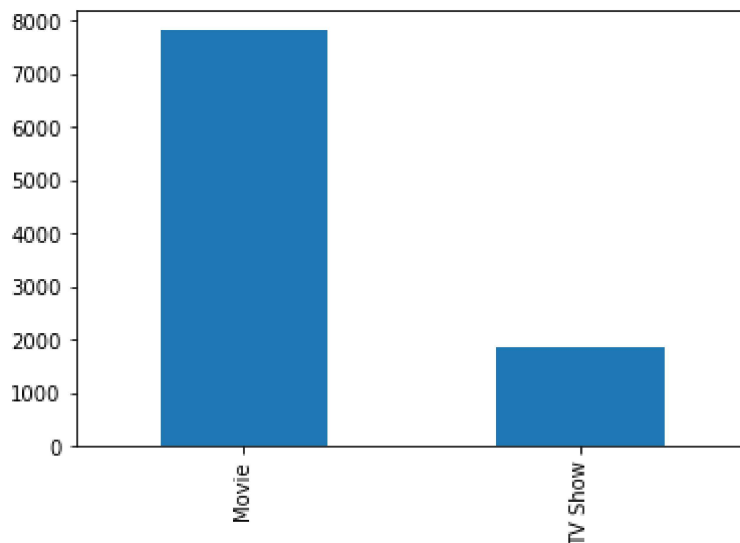
## How many Movies and Tv shows are in the dataset

```
In [17]: df.type.value_counts()
```

```
Out[17]: Movie      7814
          TV Show   1854
          Name: type, dtype: int64
```

```
In [18]: df.type.value_counts().plot(kind = "bar")
```

```
Out[18]: <AxesSubplot:>
```



## Show all records Type is "Movies" & Country is "United Kingdom".



```
In [19]: df[(df["type"] == "Movie") & (df["country"] == "United Kingdom")].head(3)
```

```
Out[19]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating
4	s5	Movie	Monster Maker	Giles Foster	Harry Dean Stanton, Kieran O'Brien, George Cos...	United Kingdom	March 30, 2021	1989	NaN
5	s6	Movie	Living With Dinosaurs	Paul Weiland	Gregory Chisholm, Juliet Stevenson, Brian Hens...	United Kingdom	March 30, 2021	1989	NaN
14	s15	Movie	Elon Musk: The Real Life Iron Man	Sonia Anderson	Elon Musk, Per Wimmer, Julie Anderson-Ankenbra...	United Kingdom	May 2, 2021	2018	NaN

**Show all records Type is "Movies" & director is "Paul Weiland".**

```
In [23]: df[(df["type"] == "Movie") & (df["director"] == "Paul Weiland")].head(3)
```

```
Out[23]:
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	dt
5	s6	Movie	Living With Dinosaurs	Paul Weiland	Gregory Chisholm, Juliet Stevenson, Brian Hens...	United Kingdom	March 30, 2021	1989	NaN	

**Show Top 3 Directors, Who gave the heighest Number of Tv shows and Movies Released on Prime Video.**

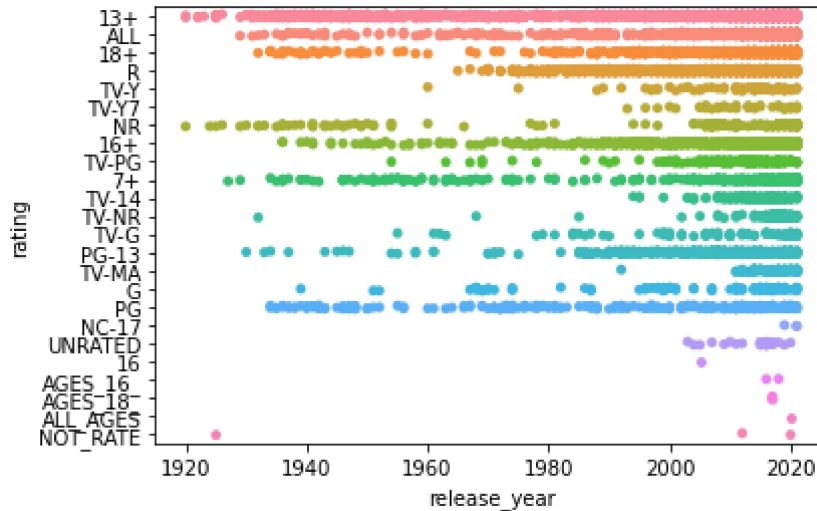
```
In [27]: df["director"].value_counts().head(3)
```

```
Out[27]: Mark Knight          113
Cannis Holder             61
Moonbug Entertainment      37
Name: director, dtype: int64
```

## in which year heighest rating show in a plot

```
In [33]: sns.stripplot(x = "release_year" , y = 'rating', data = df)
```

```
Out[33]: <AxesSubplot:xlabel='release_year', ylabel='rating'>
```



```
In [ ]:
```