project-netflix-1

June 8, 2024

[ ]:

**import warnings** warnings.filterwarnings("ignore") **import numpy as np**

**import pandas as pd**

**import matplotlib.pyplot as plt import seaborn as sns**

[ ]:

df=pd.read\_csv("https://gist.github.com/singhsidhukuldeep/

↪564f271315abb6bc22647e81e6bf4762/raw/

↪66fb67a8bb014df6b7f924aad0a91aa662bc7fc2/netflix\_titles.csv")

print(df.shape) df.head()

(8807, 12)

[ ]: 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 \

1. NaN United States
2. Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban… South Africa
3. Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi… NaN
4. NaN NaN
5. 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

[ ]:

*#Un-nesting the columns*

Pinder PinderColumns\_unnesting=["cast","country","title","director","listed\_in"]

**for** column **in** Columns\_unnesting: df[column]=df[column].str.split(', ') df=df.explode(column)

df

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

1. As her father nears the end of his life, filmm…
2. After crossing paths at a party, a Cape Town t…
3. To protect his family from a powerful drug lor…
4. Feuds, flirtations and toilet talk go down amo…
5. In a city of coaching centers known to train I…

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [ | ]: | 0 | show\_id  s1 | type Movie | Dick | Johnson | title Is Dead | director Kirsten Johnson | \ | |
|  |  | 1 | s2 | TV Show |  | Blood | & Water | NaN |  | |
|  |  | 1 | s2 | TV Show |  | Blood | & Water | NaN |  | |
|  |  | 1 | s2 | TV Show |  | Blood | & Water | NaN |  | |
|  |  | 1  …  8806 | s2  …  s8807 | TV Show  …  Movie |  | Blood  … | & Water  Zubaan | NaN  …  Mozez Singh |  | |
|  |  | 8806 | s8807 | Movie |  |  | Zubaan | Mozez Singh |  | |
|  |  | 8806 | s8807 | Movie |  |  | Zubaan | Mozez Singh |  | |
|  |  | 8806 | s8807 | Movie |  |  | Zubaan | Mozez Singh |  | |
|  |  | 8806 | s8807 | Movie |  |  | Zubaan | Mozez Singh |  | |
|  | | | cast | | | country | | date\_added | release\_year | \ |
| 0 | | | NaN | | | United States | | September 25, 2021 | 2020 |  |
| 1 | | | Ama Qamata | | | South Africa | | September 24, 2021 | 2021 |  |
| 1 | | | Ama Qamata | | | South Africa | | September 24, 2021 | 2021 |  |
| 1 | | | Ama Qamata | | | South Africa | | September 24, 2021 | 2021 |  |
| 1  …  8806 | | | Khosi Ngema  …  Anita Shabdish | | | South Africa  …  India | | September 24, 2021  …  March 2, 2019 | 2021  …  2015 |  |
| 8806 | | | Anita Shabdish | | | India | | March 2, 2019 | 2015 |  |
| 8806 | | | Chittaranjan Tripathy | | | India | | March 2, 2019 | 2015 |  |
| 8806 | | | Chittaranjan Tripathy | | | India | | March 2, 2019 | 2015 |  |
| 8806 | | | Chittaranjan Tripathy | | | India | | March 2, 2019 | 2015 |  |
| 0 | | | rating PG-13 | duration  90 min | listed\_in \ Documentaries | | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | TV-MA 2 | Seasons | International TV Shows |
| 1 | TV-MA 2 | Seasons | TV Dramas |
| 1 | TV-MA 2 | Seasons | TV Mysteries |
| 1 | TV-MA 2 | Seasons | International TV Shows |
| … | … | … | … |
| 8806 | TV-14 | 111 min | International Movies |
| 8806 | TV-14 | 111 min | Music & Musicals |
| 8806 | TV-14 | 111 min | Dramas |
| 8806 | TV-14 | 111 min | International Movies |
| 8806 | TV-14 | 111 min | Music & Musicals |

[ ]:

*#Handling null values*

Categorical\_handlingnulls = ["director","cast","country"]

**for** column **in** Categorical\_handlingnulls: df[column]=df[column].fillna(f'unknown **{**column**}**')

df

description

1. As her father nears the end of his life, filmm…
2. After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

… …

8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty…

[205237 rows x 12 columns]

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [ | ]: | 0 | show\_id  s1 | type Movie | Dick | Johnson | title Is Dead | director Kirsten Johnson | \ |
|  |  | 1 | s2 | TV Show |  | Blood | & Water | unknown director |  |
|  |  | 1 | s2 | TV Show |  | Blood | & Water | unknown director |  |
|  |  | 1 | s2 | TV Show |  | Blood | & Water | unknown director |  |
|  |  | 1  …  8806 | s2  …  s8807 | TV Show  …  Movie |  | Blood  … | & Water  Zubaan | unknown director  …  Mozez Singh |  |
|  |  | 8806 | s8807 | Movie |  |  | Zubaan | Mozez Singh |  |
|  |  | 8806 | s8807 | Movie |  |  | Zubaan | Mozez Singh |  |
|  |  | 8806 | s8807 | Movie |  |  | Zubaan | Mozez Singh |  |
|  |  | 8806 | s8807 | Movie |  |  | Zubaan | Mozez Singh |  |
| 0 | | | cast unknown cast | | | country United States | | date\_added September 25, 2021 | release\_year \  2020 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Ama Qamata | | | | South Africa | September 24, | | 2021 |  | 2021 |
| 1 | Ama Qamata | | | | South Africa | September 24, | | 2021 |  | 2021 |
| 1 | Ama Qamata | | | | South Africa | September 24, | | 2021 |  | 2021 |
| 1  …  8806 | Khosi Ngema  …  Anita Shabdish | | | | South Africa  …  India | September 24,  …  March 2, | | 2021  2019 | … | 2021  2015 |
| 8806 | Anita Shabdish | | | | India | March 2, | | 2019 |  | 2015 |
| 8806 | Chittaranjan Tripathy | | | | India | March 2, | | 2019 |  | 2015 |
| 8806 | Chittaranjan Tripathy | | | | India | March 2, | | 2019 |  | 2015 |
| 8806 | Chittaranjan Tripathy | | | | India | March 2, | | 2019 |  | 2015 |
| 0 | rating PG-13 | duration  90 min | | | listed\_in Documentaries | | \ | | | |
| 1 | TV-MA | 2 | Seasons | International TV Shows | | | | | | |
| 1 | TV-MA | 2 | Seasons | TV Dramas | | | | | | |
| 1 | TV-MA | 2 | Seasons | TV Mysteries | | | | | | |
| 1  …  8806 | TV-MA  …  TV-14 | 2 | Seasons  …  111 min | International TV Shows  …  International Movies | | | | | | |
| 8806 | TV-14 |  | 111 min | Music & Musicals | | | | | | |
| 8806 | TV-14 |  | 111 min | Dramas | | | | | | |
| 8806 | TV-14 |  | 111 min | International Movies | | | | | | |
| 8806 | TV-14 |  | 111 min | Music & Musicals | | | | | | |

description

1. As her father nears the end of his life, filmm…
2. After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

… …

8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty…

[205237 rows x 12 columns]

[ ]:

Numerical\_handlingnulls = ["duration","rating","date\_added"]

**for** c **in** Numerical\_handlingnulls: df[column]=df[column].fillna(0)

df

[ ]: show\_id type title director \

1. s1 Movie Dick Johnson Is Dead Kirsten Johnson
2. s2 TV Show Blood & Water unknown director

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 |  | s2 | TV Show | | | Blood & Water | unknown director | | | |
| 1 |  | s2 | TV Show | | | Blood & Water | unknown director | | | |
| 1  … | … | s2 | TV Show  … | | | Blood & Water  … | unknown director  … | | | |
| 8806 | s8807 | | Movie | | | Zubaan | Mozez Singh | | | |
| 8806 | s8807 | | Movie | | | Zubaan | Mozez Singh | | | |
| 8806 | s8807 | | Movie | | | Zubaan | Mozez Singh | | | |
| 8806 | s8807 | | Movie | | | Zubaan | Mozez Singh | | | |
| 8806 | s8807 | | Movie | | | Zubaan | Mozez Singh | | | |
|  | cast | | | | | country | date\_added | | release\_year | \ |
| 0 | unknown cast | | | | | United States | September 25, 2021 | | 2020 |  |
| 1 | Ama Qamata | | | | | South Africa | September 24, 2021 | | 2021 |  |
| 1 | Ama Qamata | | | | | South Africa | September 24, 2021 | | 2021 |  |
| 1 | Ama Qamata | | | | | South Africa | September 24, 2021 | | 2021 |  |
| 1  …  8806 | Khosi Ngema  …  Anita Shabdish | | | | | South Africa  …  India | September 24, 2021  …  March 2, 2019 | | 2021  …  2015 |  |
| 8806 | Anita Shabdish | | | | | India | March 2, 2019 | | 2015 |  |
| 8806 | Chittaranjan Tripathy | | | | | India | March 2, 2019 | | 2015 |  |
| 8806 | Chittaranjan Tripathy | | | | | India | March 2, 2019 | | 2015 |  |
| 8806 | Chittaranjan Tripathy | | | | | India | March 2, 2019 | | 2015 |  |
| 0 | rating PG-13 | | duration  90 min | | | listed\_in Documentaries | | \ | | |
| 1 | TV-MA | | 2 | Seasons | International TV Shows | | | | | |
| 1 | TV-MA | | 2 | Seasons | TV Dramas | | | | | |
| 1 | TV-MA | | 2 | Seasons | TV Mysteries | | | | | |
| 1  …  8806 | TV-MA  …  TV-14 | | 2 | Seasons  …  111 min | International TV Shows  …  International Movies | | | | | |
| 8806 | TV-14 | |  | 111 min | Music & Musicals | | | | | |
| 8806 | TV-14 | |  | 111 min | Dramas | | | | | |
| 8806 | TV-14 | |  | 111 min | International Movies | | | | | |
| 8806 | TV-14 | |  | 111 min | Music & Musicals | | | | | |

description

1. As her father nears the end of his life, filmm…
2. After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

1 After crossing paths at a party, a Cape Town t…

… …

8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty… 8806 A scrappy but poor boy worms his way into a ty…

[ ]:

8806 A scrappy but poor boy worms his way into a ty… [205237 rows x 12 columns]

[ ]: show\_id 0

*#Check for the null values after Handling null values.*

df.isna().sum()

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

[ ]:

*#1. Find the counts of each categorical variable both using graphical and*␣

↪*nongraphical analysis.*

*#a. For Non-graphical Analysis:*

*#Country- Column*

df.groupby("country")["title"].nunique().sort\_values(ascending=**False**)

[ ]: country

United States 3739

India 1052

unknown country 847

United Kingdom 811

Canada 450

…

Montenegro 1

Somalia 1

Mongolia 1

Botswana 1

Poland, 1

Name: title, Length: 128, dtype: int64

**INSIGHT:** By this we understood, United States, India are on the top with highest no. of movies or TV shows

[ ]:

*#Cast- Column*

df.groupby("cast")["title"].nunique().sort\_values(ascending=**False**)

[ ]: cast

unknown cast 844

Anupam Kher 43

Shah Rukh Khan 35

Naseeruddin Shah 35

Julie Tejwani 33

…

[ ]:

|  |  |
| --- | --- |
| Jane McLean | 1 |
| Jane McCarry | 1 |
| Jane Kaczmarek | 1 |
| Jane Hallaren | 1 |
| Ṣọpẹ́ Dìrísù | 1 |
| Name: title, Length: | 36440, dtype: int64 |

**INSIGHT**: Here we can see most popular actions are Unknown(null), Anupam Kher with 43 no. of movies casted followed by Shah Rukh Khan, Naseeruddin Shah.

*#Director- Column*

df.groupby("director")["title"].nunique().sort\_values(ascending=**False**)

[ ]: director

unknown director 2669

Jan Suter 22

Rajiv Chilaka 22

Raúl Campos 19

Marcus Raboy 17

…

Julio Medem 1

Julio Quintana 1

Julius Avery 1

Julius Onah 1

Barry Jenkins 1

Name: title, Length: 4994, dtype: int64

**INSIGHT**: Most of the movies or TV shows directed by unknown director & Jan Suter, Rajiv Chilaka.

[ ]:

*#Title- Column*

df.groupby("title")["title"].nunique().sort\_values(ascending=**False**)

[ ]: title

|  |  |
| --- | --- |
| #Alive | 1 |
| Rise: Ini Kalilah | 1 |
| Rip Tide | 1 |
| Riphagen - The Untouchable | 1 |
| Ripper Street | 1 |
|  | .. |
| Happy New Year | 1 |

|  |  |  |
| --- | --- | --- |
| Happy Jail |  | 1 |
| Happy Hunting |  | 1 |
| Happy Go Lucky  : | 1 | 1 |

Name: title, Length: 8927, dtype: int64

**INSIGHT**: Every title is unique here.

[ ]:

*#Type- Column*

df.groupby("type")["title"].nunique().sort\_values(ascending=**False**)

[ ]: type

Movie 6224

TV Show 2707

Name: title, dtype: int64

**INSIGHT**: By the above code we understood there are 6224 movies and 2707 TV shows.

[ ]:

*#Rating- Column*

df.groupby("rating")["title"].nunique().sort\_values(ascending=**False**)

[ ]: rating

TV-MA 3273

TV-14 2190

TV-PG 872

R 809

PG-13 494

TV-Y7 338

TV-Y 310

PG 290

TV-G 226

NR 80

G 41

TV-Y7-FV 6

0 4

NC-17 3

UR 3

66 min 1

84 min 1

74 min 1

Name: title, dtype: int64

**INSIGHT**: TV-MA, TV-14, TV-PG, R, PG-13 are the top ratings for the movies ot TV shows.

[ ]:

*#listed\_in- Column*

df.groupby("listed\_in")["title"].nunique().sort\_values(ascending=**False**)

[ ]: listed\_in

International Movies 2794

Dramas 2450

Comedies 1710

International TV Shows 1371

Documentaries 891

Action & Adventure 867

TV Dramas 770

Independent Movies 768

Children & Family Movies 648

Romantic Movies 626

TV Comedies 585

Thrillers 582

Crime TV Shows 478

Kids' TV 455

Docuseries 401

Music & Musicals 381

Romantic TV Shows 378

Stand-Up Comedy 360

Horror Movies 358

Reality TV 260

British TV Shows 255

Sci-Fi & Fantasy 243

Sports Movies 223

Spanish-Language TV Shows 180

Anime Series 177

TV Action & Adventure 169

Korean TV Shows 155

Classic Movies 116

LGBTQ Movies 105

TV Mysteries 99

Science & Nature TV 92

TV Sci-Fi & Fantasy 86

TV Horror 76

Cult Movies 72

Anime Features 71

Teen TV Shows 71

Faith & Spirituality 68

Movies 59

TV Thrillers 57

Stand-Up Comedy & Talk Shows 56

Classic & Cult TV 29

TV Shows 16

Name: title, dtype: int64

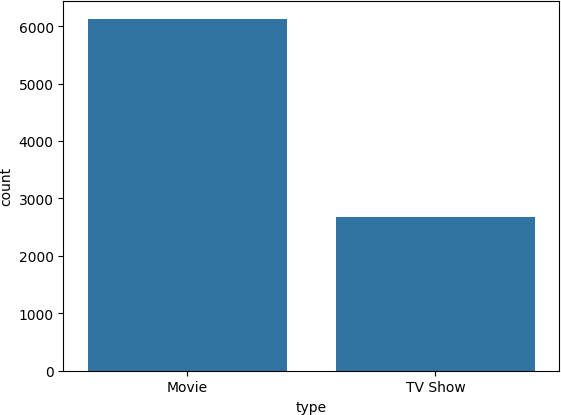
**INSIGHT**: International Movies,Dramas,Comedy, International TV shows, Documentaries are top 5 listed movies genres.

[ ]:

*#b.Non graphical analysis #type- Column*

sns.countplot(data=df, x="type") plt.show()

[ ]:

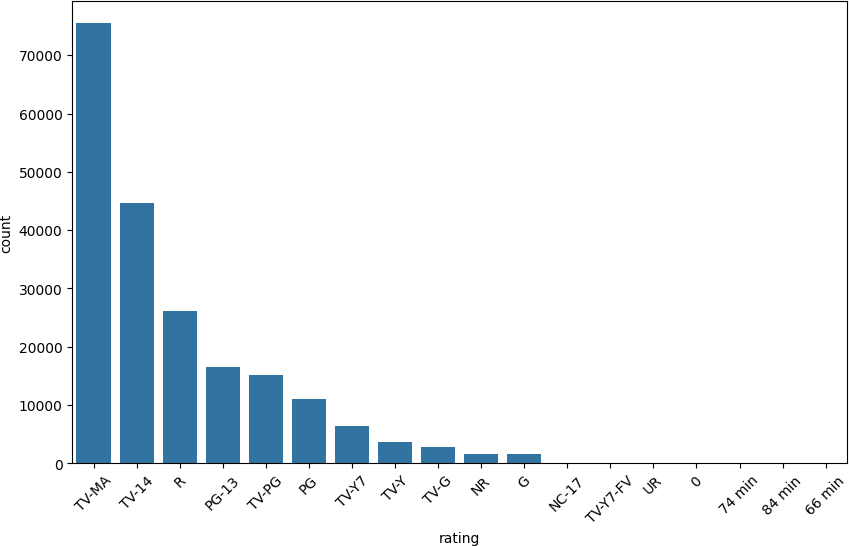


**INSIGHT**: By this graphical analysis we understood there are more number of movies than TV shows in the Netflix.

*#Rating- Column* cat\_counts=df["rating"].value\_counts() plt.figure(figsize=(10,6))

sns.countplot(data=df, x="rating",order=cat\_counts.index) plt.xticks(rotation=45)

plt.show()



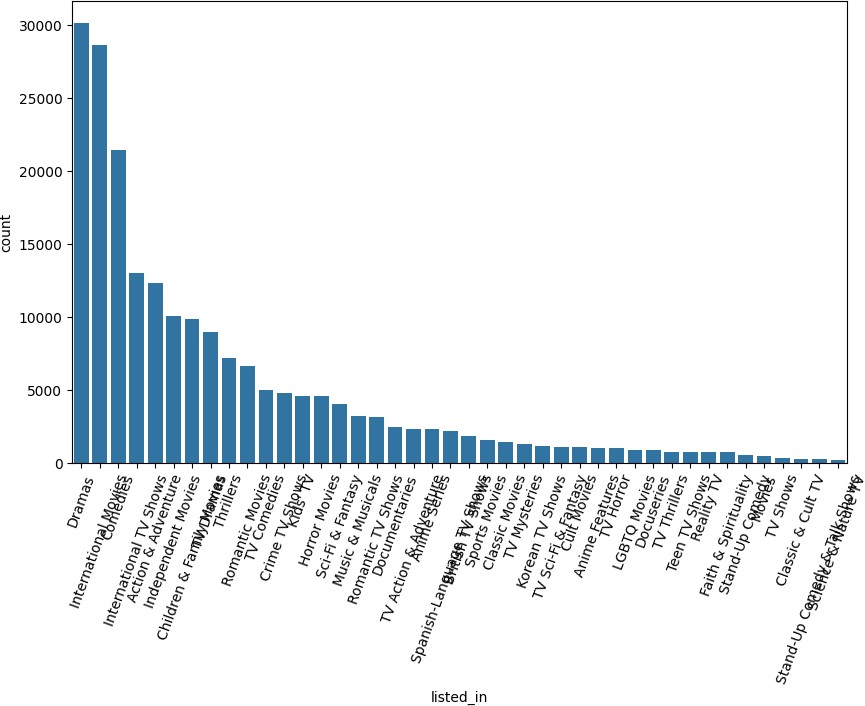
**INSIGHT**: TV-MA, TV-14, R are with the highest rating followed by PG-13, TV-PG, PG.

[ ]:

*#listed\_in- Column* cat\_counts=df["listed\_in"].value\_counts() plt.figure(figsize=(10,6))

sns.countplot(data=df, x="listed\_in",order=cat\_counts.index) plt.xticks(rotation=70)

plt.show()



[ ]:

**INSIGHT**: International movies, Dramas, Comedies are top genres that people watch on Netflix.So people are more interested in this genres.

[ ]: country

*#2. Comparison of tv shows vs movies.*

*#a. Find the number of movies produced in each country and pick the top 10*␣

↪*countries.*

df\_Movie=df.loc[df["type"]=="Movie"] df\_Movie.groupby("country")["title"].nunique().sort\_values(ascending=**False**).

↪head(10)

|  |  |
| --- | --- |
| United States | 2792 |
| India | 968 |
| United Kingdom | 538 |
| unknown country | 450 |
| Canada | 324 |
| France | 308 |
| Germany | 185 |
| Spain | 175 |

[ ]:

[ ]:

df=sns.load\_dataset('titanic') df

Japan 123

Mexico 117

Name: title, dtype: int64

**import seaborn as sns**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [ | ]: | survived | pclass | sex | age |  | sibsp parch | fare | embarked | class | \ |
|  |  | 0 0 | 3 | male | 22.0 |  | 1 0 | 7.2500 | S | Third |  |
|  |  | 1 1 | 1 | female | 38.0 |  | 1 0 | 71.2833 | C | First |  |
|  |  | 2 1 | 3 | female | 26.0 |  | 0 0 | 7.9250 | S | Third |  |
|  |  | 3 1 | 1 | female | 35.0 |  | 1 0 | 53.1000 | S | First |  |
|  |  | 4 0 | 3 | male | 35.0 |  | 0 0 | 8.0500 | S | Third |  |
|  |  | .. … | … | … … | … | … | … | … … |  |  |  |
|  |  | 886 0 | 2 | male | 27.0 |  | 0 0 | 13.0000 | S | Second |  |
|  |  | 887 1 | 1 | female | 19.0 |  | 0 0 | 30.0000 | S | First |  |
|  |  | 888 0 | 3 | female | NaN |  | 1 2 | 23.4500 | S | Third |  |
|  |  | 889 1 | 1 | male | 26.0 |  | 0 0 | 30.0000 | C | First |  |
|  |  | 890 0 | 3 | male | 32.0 |  | 0 0 | 7.7500 | Q | Third |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | who | adult\_male | deck | embark\_town | alive | alone |
| 0 | man | True | NaN | Southampton | no | False |
| 1 | woman | False | C | Cherbourg | yes | False |
| 2 | woman | False | NaN | Southampton | yes | True |
| 3 | woman | False | C | Southampton | yes | False |
| 4 | man | True | NaN | Southampton | no | True |
| .. | … | … … |  | … … … |  |  |
| 886 | man | True | NaN | Southampton | no | True |
| 887 | woman | False | B | Southampton | yes | True |
| 888 | woman | False | NaN | Southampton | no | False |
| 889 | man | True | C | Cherbourg | yes | True |
| 890 | man | True | NaN | Queenstown | no | True |

[891 rows x 15 columns]

[ ]:

**import seaborn as sns**

df=sns.load\_dataset('taxis')

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

[ ]: 44

[ ]:

*#df.groupby("embark\_town")["fare"].mean()*

df["fare"].mean()

[ ]: 32.204207968574636 [ ]:

**import pandas as pd**

pd.read\_csv('https://d2beiqkhq929f0.cloudfront.net/public\_assets/assets/000/066/

↪212/original/Ecommerce\_Purchases.txt?1708760410')

[ ]: Address Lot AM or PM \

1. 16629 Pace Camp Apt. 448\nAlexisborough, NE 77… 46 in PM
2. 9374 Jasmine Spurs Suite 508\nSouth John, TN 8… 28 rn PM
3. Unit 0065 Box 5052\nDPO AP 27450 94 vE PM
4. 7780 Julia Fords\nNew Stacy, WA 45798 36 vm PM
5. 23012 Munoz Drive Suite 337\nNew Cynthia, TX 5… 20 IE AM

… … … …

9995 966 Castaneda Locks\nWest Juliafurt, CO 96415 92 XI PM 9996 832 Curtis Dam Suite 785\nNorth Edwardburgh, T… 41 JY AM

|  |  |  |  |
| --- | --- | --- | --- |
| 9997 | Unit 4434 Box 6343\nDPO AE 28026-0283 | 74 Zh | AM |
| 9998 | 0096 English Rest\nRoystad, IA 12457 | 74 cL | PM |
| 9999 | 40674 Barrett Stravenue\nGrimesville, WI 79682 | 64 Hr | AM |
| 0 | Browser Info Opera/9.56.(X11; Linux x86\_64; sl-SI) Presto/2… | \ |  |
| 1 | Opera/8.93.(Windows 98; Win 9x 4.90; en-US) Pr… |  |  |
| 2 | Mozilla/5.0 (compatible; MSIE 9.0; Windows NT … |  |  |
| 3 | Mozilla/5.0 (Macintosh; Intel Mac OS X 10\_8\_0 … |  |  |
| 4  …  9995 | Opera/9.58.(X11; Linux x86\_64; it-IT) Presto/2…  …  Mozilla/5.0 (Windows NT 5.1) AppleWebKit/5352 … |  |  |
| 9996 | Mozilla/5.0 (compatible; MSIE 9.0; Windows NT … |  |  |
| 9997 | Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10\_7… |  |  |
| 9998 | Mozilla/5.0 (Macintosh; Intel Mac OS X 10\_8\_8;… |  |  |
| 9999 | Mozilla/5.0 (X11; Linux i686; rv:1.9.5.20) Gec… |  |  |
|  | Company Credit Card | CC Exp Date | \ |
| 0 | Martinez-Herman 6011929061123406 | 02/20 |  |
| 1 | Fletcher, Richards and Whitaker 3337758169645356 | 11/18 |  |
| 2 | Simpson, Williams and Pham 675957666125 | 08/19 |  |
| 3 | Williams, Marshall and Buchanan 6011578504430710 | 02/24 |  |
| 4  …  9995 | Brown, Watson and Andrews 6011456623207998  … …  Randall-Sloan 342945015358701 | 10/25  …  03/22 |  |
| 9996 | Hale, Collins and Wilson 210033169205009 | 07/25 |  |
| 9997 | Anderson Ltd 6011539787356311 | 05/21 |  |
| 9998 | Cook Inc 180003348082930 | 11/17 |  |
| 9999 | Greene Inc 4139972901927273 | 02/19 |  |
| 0 | CC Security Code CC Provider \ 900 JCB 16 digit |  |  |

1. 561 Mastercard
2. 699 JCB 16 digit
3. 384 Discover
4. 678 Diners Club / Carte Blanche

… … …

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 9995 | 838 | JCB 15 digit | | | | | |
| 9996 | 207 | JCB 16 digit | | | | | |
| 9997 | 1 | VISA 16 digit | | | | | |
| 9998 | 987 | American Express | | | | | |
| 9999 | 302 | JCB 15 digit | | | | | |
| 0 | Email [pdunlap@yahoo.com](mailto:pdunlap@yahoo.com) | | | Scientist, | | Job product/process development | \ |
| 1 | [anthony41@reed.com](mailto:anthony41@reed.com) | | |  | | Drilling engineer |  |
| 2 | [amymiller@morales-harrison.com](mailto:amymiller@morales-harrison.com) | | |  | | Customer service manager |  |
| 3 | [brent16@olson-robinson.info](mailto:brent16@olson-robinson.info) | | |  | | Drilling engineer |  |
| 4  … 9995 | [christopherwright@gmail.com](mailto:christopherwright@gmail.com)  … [iscott@wade-garner.com](mailto:iscott@wade-garner.com) | | |  | | Fine artist  … Printmaker |  |
| 9996 | [mary85@hotmail.com](mailto:mary85@hotmail.com) | | |  | | Energy engineer |  |
| 9997 | [tyler16@gmail.com](mailto:tyler16@gmail.com) | | |  | | Veterinary surgeon |  |
| 9998 | [elizabethmoore@reid.net](mailto:elizabethmoore@reid.net) | | |  | | Local government officer |  |
| 9999 | [rachelford@vaughn.com](mailto:rachelford@vaughn.com) | | |  | | Embryologist, clinical |  |
|  | IP Address | Language | Purchase | | Price | | |
| 0 | 149.146.147.205 | el |  | | 98.14 | | |
| 1 | 15.160.41.51 | fr |  | | 70.73 | | |
| 2 | 132.207.160.22 | de |  | | 0.95 | | |
| 3 | 30.250.74.19 | es |  | | 78.04 | | |
| 4  … 9995 | 24.140.33.94  … 29.73.197.114 | es  …  it | … | | 77.82  82.21 | | |
| 9996 | 121.133.168.51 | pt |  | | 25.63 | | |
| 9997 | 156.210.0.254 | el |  | | 83.98 | | |
| 9998 | 55.78.26.143 | es |  | | 38.84 | | |
| 9999 | 176.119.198.199 | el |  | | 67.59 | | |

[10000 rows x 14 columns]

[ ]:

*#b. Find the number of Tv-Shows produced in each country and pick the top 10*␣

↪*countries.*

df\_tvshow=df.loc[df["type"]=="TV Show"] df\_tvshow.groupby("country")["title"].nunique().sort\_values(ascending=**False**).

↪head(10)

[ ]: country

United States 950

|  |  |
| --- | --- |
| unknown country | 397 |
| United Kingdom | 273 |
| Japan | 200 |
| South Korea | 174 |
| Canada | 126 |
| France | 90 |
| India | 84 |
| Taiwan | 70 |
| Australia | 66 |

[ ]:

Name: title, dtype: int64

**INSIGHT**: Here we can see United States is the only Country which is top in movies and TV shows and other countries like India is in second place in Movies while Unknown Country in TV shows. It is to recommend that US to continue producing and get many other from other countries as well.

*#3. What is the best time to launch a TV show?*

*#a. Find which is the best week to release the Tv-show or the movie. Do the*␣

↪*analysis separately for Tv-shows and Movies #Convert date\_added to datetime format*

df['date\_added']=pd.to\_datetime(df['date\_added'],format='mixed')

[ ]:

*#extracting week,month,year from the date\_added column*

df["month\_added"]=df["date\_added"].dt.month

[ ]:

df["year\_added"]=df["date\_added"].dt.year

[ ]:

df["week\_added"]=df["date\_added"].dt.strftime('%U')

[ ]:

df\_tvshow=df.loc[df["type"]=="TV Show"]

[ ]:

df\_tvshow.groupby('week\_added')['title'].nunique().sort\_values(ascending=**False**).

↪head().reset\_index(name="No\_of\_shows\_released")

[ ]: week\_added No\_of\_shows\_released

|  |  |  |
| --- | --- | --- |
| 0 | 39 | 96 |
| 1 | 31 | 87 |
| 2 | 13 | 85 |
| 3 | 26 | 84 |
| 4 | 27 | 83 |

[ ]:

df\_Movie=df.loc[df["type"]=="Movie"]

[ ]:

df\_Movie.groupby('week\_added')['title'].nunique().sort\_values(ascending=**False**).

↪head().reset\_index(name="No\_of\_movies\_released")

[ ]: week\_added No\_of\_movies\_released

|  |  |  |
| --- | --- | --- |
| 0 | 00 | 255 |
| 1 | 39 | 245 |
| 2 | 26 | 236 |
| 3 | 13 | 232 |
| 4 | 43 | 207 |

[ ]:

*#b. Find which is the best month to release the Tv-show or the movie. Do the*␣

↪*analysis separately for Tv-shows and Movies*

df\_tvshow.groupby('month\_added')['title'].nunique().

↪sort\_values(ascending=**False**).head().reset\_index(name="No\_of\_shows\_released")

**INSIGHT**: By this analysis, we understood on average 39th, 26th, 13th weeks are the best time to release a movie or TV show in the Netflix.

|  |  |  |  |
| --- | --- | --- | --- |
| [ | ]: | month\_added | No\_of\_shows\_released |
|  |  | 0 12 | 269 |
|  |  | 1 7 | 264 |
|  |  | 2 9 | 253 |
|  |  | 3 6 | 239 |
|  |  | 4 8 | 237 |

[ ]:

df\_Movie.groupby('month\_added')['title'].nunique().sort\_values(ascending=**False**).

↪head().reset\_index(name="No\_of\_movies\_released")

|  |  |  |  |
| --- | --- | --- | --- |
| [ | ]: | month\_added | No\_of\_movies\_released |
|  |  | 0 7 | 572 |
|  |  | 1 12 | 559 |
|  |  | 2 4 | 555 |
|  |  | 3 1 | 554 |
|  |  | 4 10 | 554 |

[ ]:

**INSIGHT**: By this analysis, we understood on average 7th, 12th months are the best time to release a movie or TV show in the Netflix. Approx it is mid of the year and end of the year when the movie or tv show released in Netflix.

[ ]: cast No\_of\_TV Shows/Movies

*#4. Analysis of actors/directors of different types of shows/movies.*

*#a. Identify the top 10 directors who have appeared in most movies or TV shows.*

df.groupby("cast")["title"].nunique().sort\_values(ascending=**False**).head(10).

↪reset\_index(name='No\_of\_TV Shows/Movies')

|  |  |
| --- | --- |
| 0 unknown cast | 844 |
| 1 Anupam Kher | 43 |
| 2 Shah Rukh Khan | 35 |
| 3 Naseeruddin Shah | 35 |

|  |  |
| --- | --- |
| 4 Julie Tejwani | 33 |
| 5 Takahiro Sakurai | 32 |
| 6 Rupa Bhimani | 31 |
| 7 Om Puri | 30 |
| 8 Akshay Kumar | 30 |
| 9 Yuki Kaji | 29 |

[ ]:

*#b. Identify the top 10 directors who have appeared in most movies or TV shows.*

df.groupby("director")["title"].nunique().sort\_values(ascending=**False**).head(10).

↪reset\_index(name='No\_of\_TV Shows/Movies')

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| [ | ]: |  | director | No\_of\_TV Shows/Movies |
|  |  | 0 | unknown director | 2669 |
|  |  | 1 | Jan Suter | 22 |
|  |  | 2 | Rajiv Chilaka | 22 |
|  |  | 3 | Raúl Campos | 19 |
|  |  | 4 | Marcus Raboy | 17 |
|  |  | 5 | Suhas Kadav | 16 |
|  |  | 6 | Cathy Garcia-Molina | 15 |
|  |  | 7 | Jay Karas | 15 |
|  |  | 8 | Jay Chapman | 12 |
|  |  | 9 | Martin Scorsese | 12 |

[ ]:

*#5. Which genre movies are more popular or produced more*

**from wordcloud import** WordCloud text=' '.join(df["listed\_in"])

wordcloud=WordCloud(width=700,height=300,background\_color='yellow').

↪generate(text)

plt.figure(figsize=(12,8)) plt.imshow(wordcloud,interpolation='lanczos') plt.axis('off')

plt.show()

**INSIGHT**: We can identify top 10 directors & actors who have appeared in most movies & TV shows by the above codes.At the top Unknown actor & unknown Director is placed.



[ ]:

**INSIGHT**: By the plot, we can observe most popular genre in Netflix is International Movies followed by Tv shows, Comdey, Dramas. So these kind of genres are most recommended to the Netflix to get more popular.

*#6. Find After how many days the movie will be added to Netflix after the*␣

↪*release of the movie.*

df["year\_diff"]=df["year\_added"]-df["release\_year"] df["year\_diff"]

|  |  |  |  |
| --- | --- | --- | --- |
| [ | ]: | 0 | 1 |
|  |  | 1 | 0 |
|  |  | 1 | 0 |
|  |  | 1 | 0 |
|  |  | 1 | 0 |
|  |  |  | .. |
|  |  | 8806 | 4 |
|  |  | 8806 | 4 |
|  |  | 8806 | 4 |
|  |  | 8806 | 4 |
|  |  | 8806 | 4 |
|  |  | Name: | year\_diff, Length: 205237, dtype: int64 |

[ ]:

df["year\_diff"].mode()

[ ]: 0 0

Name: year\_diff, dtype: int64

**INSIGHT**: With the mode() function understood that every day a movie will be released in the Netflix. For better future growth Netflix should continue this.