Amazon Prime Notebook

June 21, 2024

1 Uploading Necessary Modules

```python
[3]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
```

1.1 Reading the CSV file

```python
[68]: df=pd.read_csv("amazon_prime_titles.csv")
```

```python
[69]: df.head()
```

```
[69]:   show_id   type                title         director  \
     0      s1  Movie    The Grand Seduction    Don McKellar
     1      s2  Movie    Take Care Good Night    Girish Joshi
     2      s3  Movie    Secrets of Deception     Josh Webber
     3      s4  Movie       Pink: Staying True  Sonia Anderson
     4      s5  Movie            Monster Maker    Giles Foster

                                          cast          country  \
     0      Brendan Gleeson, Taylor Kitsch, Gordon Pinsent          Canada
     1   Mahesh Manjrekar, Abhay Mahajan, Sachin Khedekar          India
     2  Tom Sizemore, Lorenzo Lamas, Robert LaSardo, R…   United States
     3  Interviews with: Pink, Adele, Beyoncé, Britney…   United States
     4  Harry Dean Stanton, Kieran O'Brien, George Cos…  United Kingdom

            date_added  release_year rating duration              listed_in  \
     0  March 30, 2021          2014    NaN  113 min           Comedy, Drama
     1  March 30, 2021          2018    13+  110 min     Drama, International
     2  March 30, 2021          2017    NaN   74 min  Action, Drama, Suspense
     3  March 30, 2021          2014    NaN   69 min             Documentary
     4  March 30, 2021          1989    NaN   45 min          Drama, Fantasy

                                       description
     0  A small fishing village must procure a local d…
     1  A Metro Family decides to fight a Cyber Crimin…
```

```
2  After a man discovers his wife is cheating on …
3  Pink breaks the mold once again, bringing her …
4  Teenage Matt Banting wants to work with a famo…
```

[7]: `df.tail()`

[7]:
```
      show_id     type                          title          director  \
9663    s9664    Movie          Pride Of The Bowery  Joseph H. Lewis
9664    s9665  TV Show                Planet Patrol               NaN
9665    s9666    Movie                      Outpost      Steve Barker
9666    s9667  TV Show      Maradona: Blessed Dream               NaN
9667    s9668    Movie                   Harry Brown      Daniel Barber

                                                 cast country date_added  \
9663                          Leo Gorcey, Bobby Jordan     NaN        NaN
9664  DICK VOSBURGH, RONNIE STEVENS, LIBBY MORRIS, M…     NaN        NaN
9665  Ray Stevenson, Julian Wadham, Richard Brake, M…     NaN        NaN
9666  Esteban Recagno, Ezequiel Stremiz, Luciano Vit…     NaN        NaN
9667  Michael Caine, Emily Mortimer, Joseph Gilgun, …     NaN        NaN

      release_year rating   duration                 listed_in  \
9663          1940     7+     60 min                    Comedy
9664          2018    13+  4 Seasons                  TV Shows
9665          2008      R     90 min                    Action
9666          2021  TV-MA   1 Season              Drama, Sports
9667          2010      R    103 min  Action, Drama, Suspense

                                            description
9663  New York City street principles get an East Si…
9664  This is Earth, 2100AD – and these are the adve…
9665  In war-torn Eastern Europe, a world-weary grou…
9666  The series tells the story of Diego Maradona, …
9667  Harry Brown, starring two-time Academy Award w…
```

[12]: `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    7585 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
```

1.2   Baisc Statistics of the Data Set

```
[16]: df.describe()
```

```
[16]:         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
```

```
[18]: df.size
```

```
[18]: 116016
```

```
[20]: df.shape
```

```
[20]: (9668, 12)
```

```
[158]: df.columns
```

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

1.3   Calculating Duplicates

```
[70]: df.duplicated()
```

```
[70]: 0       False
      1       False
      2       False
      3       False
      4       False
              …
      9663    False
      9664    False
```

```
9665    False
9666    False
9667    False
Length: 9668, dtype: bool
```

1.4 Finding the Director who did more movies and Films

```
[88]: df["director"].value_counts().head(5).to_frame()
```

```
[88]:                         count
      director
      Mark Knight               113
      Cannis Holder              61
      Moonbug Entertainment      37
      Jay Chapman                34
      Arthur van Merwijk         30
```

1.5 Finding out that which countries directors are more in the set

```
[89]: df["country"].value_counts().head(3).to_frame()
```

```
[89]:                  count
      country
      United States      253
      India              229
      United Kingdom      28
```

1.6 Dropping the Show_id column since it is un-necessary

```
[88]: df.head()
```

```
[88]:     type              title          director  \
      0  Movie    The Grand Seduction      Don McKellar
      1  Movie    Take Care Good Night     Girish Joshi
      2  Movie    Secrets of Deception      Josh Webber
      3  Movie      Pink: Staying True   Sonia Anderson
      4  Movie            Monster Maker     Giles Foster

                                                     cast          country  \
      0     Brendan Gleeson, Taylor Kitsch, Gordon Pinsent           Canada
      1   Mahesh Manjrekar, Abhay Mahajan, Sachin Khedekar           India
      2  Tom Sizemore, Lorenzo Lamas, Robert LaSardo, R…   United States
      3  Interviews with: Pink, Adele, Beyoncé, Britney…   United States
      4  Harry Dean Stanton, Kieran O'Brien, George Cos…  United Kingdom

        date_added  release_year rating duration            listed_in  \
      0 2021-03-30          2014    NaN      113        Comedy, Drama
```

```
1 2021-03-30            2018    13+        110        Drama, International
2 2021-03-30            2017    NaN         74   Action, Drama, Suspense
3 2021-03-30            2014    NaN         69                Documentary
4 2021-03-30            1989    NaN         45             Drama, Fantasy

                                          description
0  A small fishing village must procure a local d…
1  A Metro Family decides to fight a Cyber Crimin…
2  After a man discovers his wife is cheating on …
3  Pink breaks the mold once again, bringing her …
4  Teenage Matt Banting wants to work with a famo…
```

[47]: `df["type"].unique()`
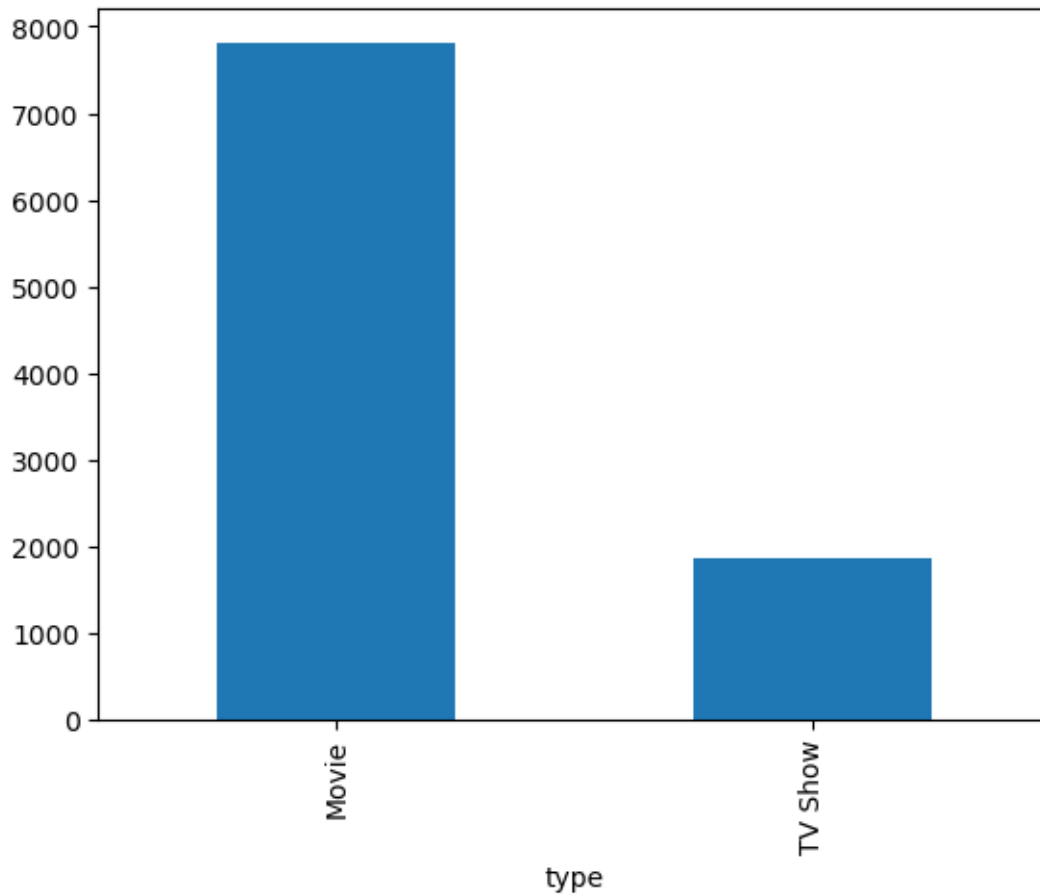
[47]: `array(['Movie', 'TV Show'], dtype=object)`

1.7 Finding the count of Movies and Tv show in Data set

[51]: `df["type"].value_counts()`

[51]:
```
type
Movie      7814
TV Show    1854
Name: count, dtype: int64
```

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

[57]: `<Axes: xlabel='type'>`

1.8 Calculation Null in Data Frame

```
[77]: df1=df.isnull().sum()
      df1.to_frame()
```

```
[77]:                    0
      type               0
      title              0
      director        2083
      cast            1233
      country         8996
      date_added      9513
      release_year       0
      rating           337
      duration           0
      listed_in           0
      description         0
```

```
[ ]: ## Filling Backward to Remove the Null values in the data set
```

```
[51]: df["rating"].bfill().tail().to_frame()
```

```
[51]:       rating
      9663      7+
      9664     13+
      9665       R
      9666   TV-MA
      9667       R
```

1.9 Displaying First 5 Rows having type "Movies" and Country "USA"

```
[84]: filter=df[(df["type"]=="Movie") & (df["country"]=="United Kingdom")]

      filter.head(5)
```

```
[84]:        type                              title           director  \
      4     Movie                     Monster Maker       Giles Foster
      5     Movie               Living With Dinosaurs       Paul Weiland
      14    Movie  Elon Musk: The Real Life Iron Man   Sonia Anderson
      374   Movie                    The Zombie King   Aidan Belizaire
      656   Movie                          The Flaw     David Sington

                                               cast          country  \
      4     Harry Dean Stanton, Kieran O'Brien, George Cos…  United Kingdom
      5     Gregory Chisholm, Juliet Stevenson, Brian Hens…  United Kingdom
      14    Elon Musk, Per Wimmer, Julie Anderson-Ankenbra…  United Kingdom
      374     Edward Furlong, Corey Feldman, George McCluskey  United Kingdom
      656           Andrew Luan, Robert Shiller, Louis Hyman  United Kingdom

                date_added  release_year rating duration          listed_in  \
      4     March 30, 2021          1989    NaN   45 min      Drama, Fantasy
      5     March 30, 2021          1989    NaN   52 min      Fantasy, Kids
      14        May 2, 2021          2018    NaN   74 min         Documentary
      374             NaN          2013    16+   96 min      Comedy, Horror
      656             NaN          2011    13+   81 min   Special Interest

                                          description
      4     Teenage Matt Banting wants to work with a famo…
      5     The story unfolds in a an English seaside town…
      14    Discover the meteoric rise of Elon Musk, the m…
      374   After heartbroken widower Samuel makes a deal …
      656   The definitive doc on the 2008 financial crash…
```

1.10 Finding is there any duplicates

```
[144]: df1=df.duplicated()

       df2=pd.DataFrame(df1)
       df2.value_counts()   #No duplicates in the Data Set
```

```
[144]: False    9668
       Name: count, dtype: int64
```

```
[150]: df["duration"].head(3)
```

```
[150]: 0    113 min
       1    110 min
       2     74 min
       Name: duration, dtype: object
```

1.11 Removing the "min" string after integer

```
[72]: df1=df['duration'] = df['duration'].replace(r'[^0-9]+', '', regex=True)
      df1.head(3)
```

```
[72]: 0    113
      1    110
      2     74
      Name: duration, dtype: object
```

```
[74]: df["date_added"]=pd.to_datetime(df["date_added"])
```

1.12 Calculating Null values in the Columns

```
[184]: for i in df.columns:
           count=df[i].isnull().sum()
           print(f"{i}:\t\t\t{count},Null values in Column {i}")
```

```
type:                 0,Null values in Column type
title:                0,Null values in Column title
director:             2083,Null values in Column director
cast:                 1233,Null values in Column cast
country:              8996,Null values in Column country
date_added:           9513,Null values in Column date_added
release_year:         0,Null values in Column release_year
rating:               337,Null values in Column rating
duration:             0,Null values in Column duration
listed_in:            0,Null values in Column listed_in
description:          0,Null values in Column description
```
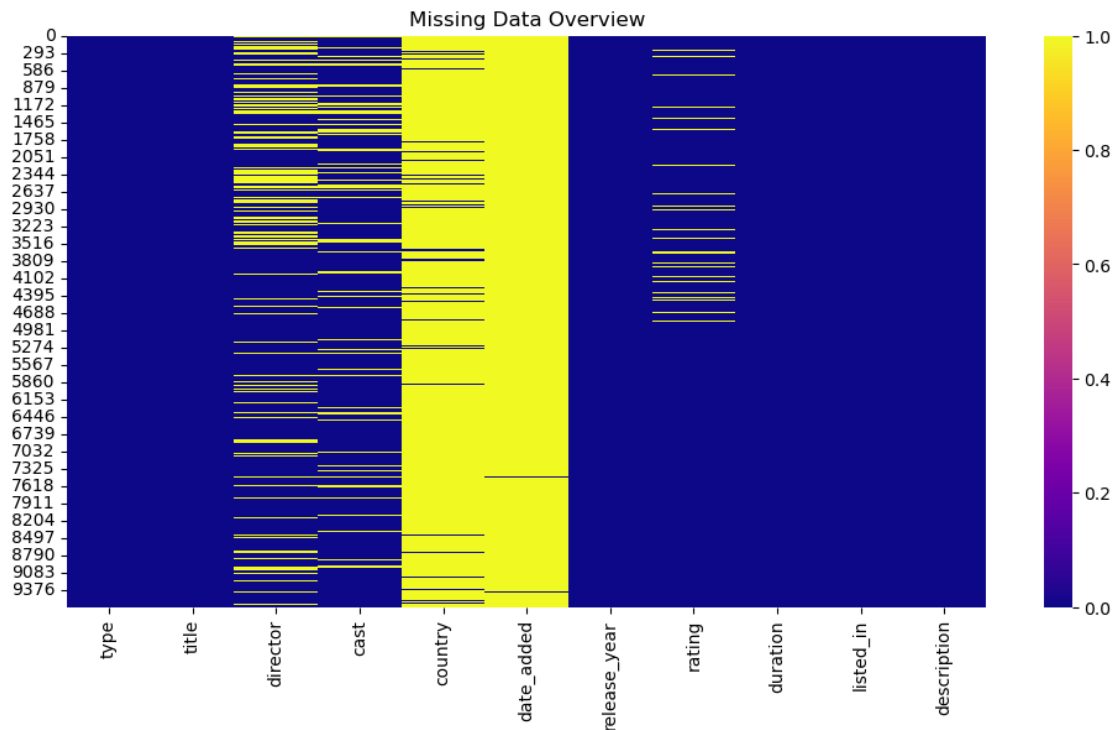
1.13 Plotting the Null values in the Data Set

```
[193]: figure,axes=plt.subplots(figsize=(12,6))

       sns.heatmap(df.isnull(),cmap="plasma",cbar=True)

       plt.title("Missing Data Overview")

       plt.show()
```


Missing Data Overview

```
[204]: df["release_year"].replace(np.nan,0)
```

```
[204]: 0        2014
       1        2018
       2        2017
       3        2014
       4        1989
                 ...
       9663     1940
       9664     2018
       9665     2008
       9666     2021
       9667     2010
       Name: release_year, Length: 9668, dtype: int64
```
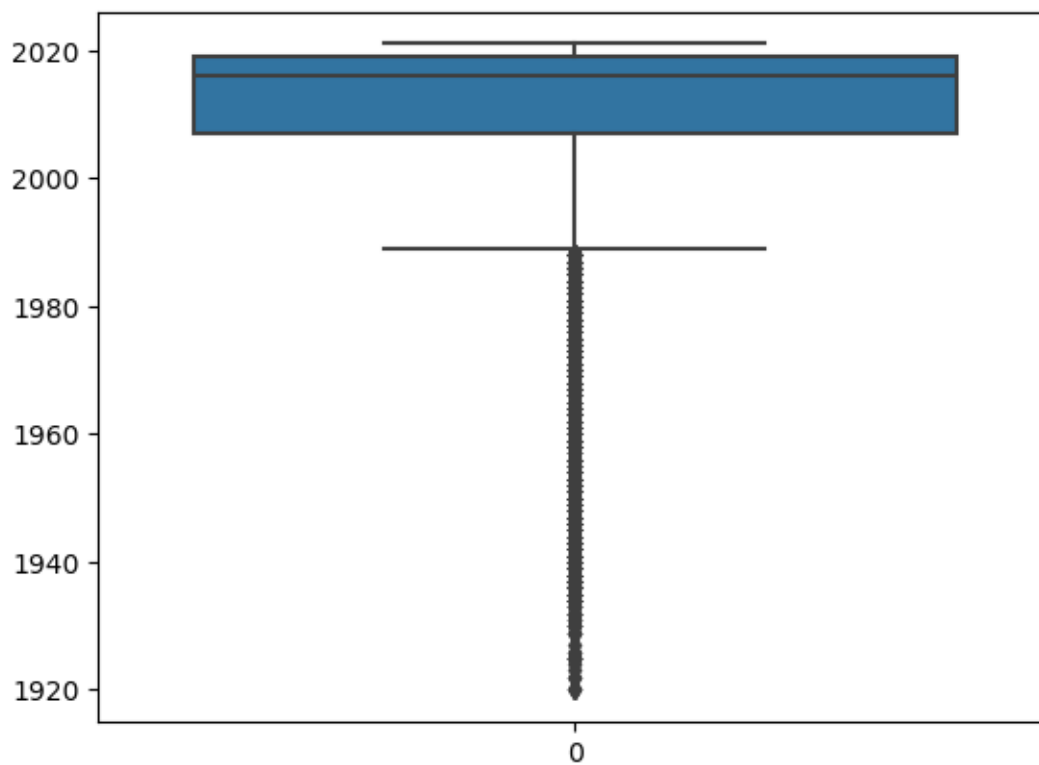
```
[257]: df["release_year"].describe()
```

```
[257]: count    9668.000000
       mean     2008.341849
       std        18.922482
       min      1920.000000
       25%      2007.000000
       50%      2016.000000
       75%      2019.000000
       max      2021.000000
       Name: release_year, dtype: float64
```

```
[258]: sns.boxplot(df["release_year"])
```

```
[258]: <Axes: >
```



```
[24]: df.head()
```

```
[24]:      type              title          director  \
      0  Movie   The Grand Seduction    Don McKellar
      1  Movie   Take Care Good Night    Girish Joshi
      2  Movie   Secrets of Deception    Josh Webber
```

```
3  Movie      Pink: Staying True  Sonia Anderson
4  Movie          Monster Maker    Giles Foster

                                         cast          country  \
0      Brendan Gleeson, Taylor Kitsch, Gordon Pinsent          Canada
1   Mahesh Manjrekar, Abhay Mahajan, Sachin Khedekar            India
2  Tom Sizemore, Lorenzo Lamas, Robert LaSardo, R…   United States
3  Interviews with: Pink, Adele, Beyoncé, Britney…   United States
4  Harry Dean Stanton, Kieran O'Brien, George Cos…  United Kingdom

        date_added  release_year rating duration             listed_in  \
0  March 30, 2021          2014    NaN  113 min           Comedy, Drama
1  March 30, 2021          2018    13+  110 min      Drama, International
2  March 30, 2021          2017    NaN   74 min   Action, Drama, Suspense
3  March 30, 2021          2014    NaN   69 min               Documentary
4  March 30, 2021          1989    NaN   45 min             Drama, Fantasy

                                         description
0  A small fishing village must procure a local d…
1  A Metro Family decides to fight a Cyber Crimin…
2  After a man discovers his wife is cheating on …
3  Pink breaks the mold once again, bringing her …
4  Teenage Matt Banting wants to work with a famo…
```

```python
[92]: mapping={

    "Movie":1,
    "TV Show":2
}

df["type"]=df["type"].map(mapping)
```

```python
[94]: df.head()
```

```
[94]:   type                title        director  \
      0     1   The Grand Seduction    Don McKellar
      1     1   Take Care Good Night    Girish Joshi
      2     1   Secrets of Deception     Josh Webber
      3     1      Pink: Staying True  Sonia Anderson
      4     1          Monster Maker    Giles Foster

                                         cast          country  \
0      Brendan Gleeson, Taylor Kitsch, Gordon Pinsent          Canada
1   Mahesh Manjrekar, Abhay Mahajan, Sachin Khedekar            India
2  Tom Sizemore, Lorenzo Lamas, Robert LaSardo, R…   United States
3  Interviews with: Pink, Adele, Beyoncé, Britney…   United States
4  Harry Dean Stanton, Kieran O'Brien, George Cos…  United Kingdom
```

```
   date_added  release_year rating duration                 listed_in  \
0  2021-03-30          2014    NaN      113            Comedy, Drama
1  2021-03-30          2018    13+      110       Drama, International
2  2021-03-30          2017    NaN       74  Action, Drama, Suspense
3  2021-03-30          2014    NaN       69                Documentary
4  2021-03-30          1989    NaN       45            Drama, Fantasy

                                          description
0  A small fishing village must procure a local d…
1  A Metro Family decides to fight a Cyber Crimin…
2  After a man discovers his wife is cheating on …
3  Pink breaks the mold once again, bringing her …
4  Teenage Matt Banting wants to work with a famo…
```

1.14 Finding all the Movies and Films in 113 min

```
[102]: filtered_data = df[(df["duration"] == "113") & (df["type"].isin([1, 2]))]
       filtered_data.head(6)
```

```
[102]:       type                      title            director  \
       0        1        The Grand Seduction       Don McKellar
       144      1  We Need to Talk About Kevin        Lynne Ramsay
       184      1                    Vinci Da    Srijit Mukherjee
       443      1       The Snows of Kilimanjaro          Henry King
       603      1                The Informer   Andrea Di Stefano
       1424     1                One Great Love          Eric Quizon

                                                    cast country date_added  \
       0          Brendan Gleeson, Taylor Kitsch, Gordon Pinsent  Canada 2021-03-30
       144              Tilda Swinton, John C. Reilly, Ezra Miller     NaN        NaT
       184    Rudranil Ghosh, Ritwick Chakraborty, Sohini Sa…     NaN        NaT
       443    Gregory Peck, Susan Hayward, Ava Gardner, Hild…     NaN        NaT
       603    Rosamund Pike, Joel Kinnamen, Common, Ana De A…     NaN        NaT
       1424   Dennis Trillo, Kim Chiu, JC de Vera, Eric Quizon     NaN        NaT

             release_year rating duration                 listed_in  \
       0              2014    NaN      113            Comedy, Drama
       144            2012      R      113  Drama, Horror, Suspense
       184            2019    16+      113  International, Suspense
       443            1952    13+      113            Action, Drama
       603            2020      R      113         Action, Suspense
       1424           2018    13+      113                    Drama

                                              description
       0        A small fishing village must procure a local d…
       144      In this Golden Globe nominated psychological t…
```

```
184     At the time when Vinci Da, an extremely talent…
443     The story of a writer of renown, lying injured…
603     An ex-convict working undercover intentionally…
1424    The story revolves around Zyra Paez, whose rel…
```

[103]:
```python
new_type=df["type"]

new_type.head(5)
```

[103]:
```
0    1
1    1
2    1
3    1
4    1
Name: type, dtype: int64
```

[115]:
```python
df["duration"]=new_duration
```

[116]:
```python
df["duration"]
```

[116]:
```
0       113
1       110
2        74
3        69
4        45
         …
9663     60
9664      4
9665     90
9666      1
9667    103
Name: duration, Length: 9668, dtype: object
```