```
import pandas as pd
df = pd.read csv("Netflix movies and tv shows clustering.csv")
print(df.info())
print(df.head())
<class 'pandas.core.frame.DataFrame'>
Index: 4808 entries, 1 to 7783
Data columns (total 12 columns):
                   Non-Null Count
#
     Column
                                    Dtype
- - -
                   _____
     _ _ _ _ _
                   4808 non-null
 0
     show id
                                    object
 1
                   4808 non-null
     type
                                    object
 2
     title
                   4808 non-null
                                    object
 3
     director
                   4808 non-null
                                    object
 4
                   4808 non-null
                                    object
     cast
 5
                   4808 non-null
     country
                                    object
 6
     date added
                   4808 non-null
                                    object
 7
     release year
                   4808 non-null
                                    int64
 8
                   4808 non-null
                                    object
     rating
 9
     duration
                   4808 non-null
                                    object
10
    listed in
                   4808 non-null
                                    object
 11
     description
                   4808 non-null
                                    object
dtypes: int64(1), object(11)
memory usage: 488.3+ KB
None
  show id
              type title
                                     director \
1
                    7:19
                           jorge michel grau
       s2
             movie
2
                    23:59
                                 gilbert chan
       s3
             movie
3
       s4
             movie
                        9
                                  shane acker
4
       s5
             movie
                       21
                               robert luketic
5
       s6 tv show
                       46
                                  serdar akar
                                                 cast
                                                             country \
   demián bichir, héctor bonilla, oscar serrano, ...
                                                               mexico
   tedd chan, stella chung, henley hii, lawrence ...
                                                           singapore
   elijah wood, john c. reilly, jennifer connelly...
                                                       united states
   jim sturgess, kevin spacey, kate bosworth, aar...
                                                       united states
  erdal beşikçioğlu, yasemin allen, melis birkan...
                                                               turkey
          date added
                      release year rating
                                            duration \
   december 23, 2016
                                              93 min
1
                               2016 tv-ma
   december 20, 2018
                               2011
                                              78 min
                                         r
3
   november 16, 2017
                               2009
                                     pq - 13
                                              80 min
4
     january 1, 2020
                              2008
                                     pq-13
                                             123 min
        july 1, 2017
                              2016
                                    tv-ma 1 season
                                            listed in \
```

```
1
                        dramas, international movies
2
                 horror movies, international movies
3
   action & adventure, independent movies, sci-fi...
4
5
     international tv shows, tv dramas, tv mysteries
                                          description
  after a devastating earthquake hits mexico cit...
  when an army recruit is found dead, his fellow...
  in a postapocalyptic world, rag-doll robots hi...
   a brilliant group of students become card-coun...
   a genetics professor experiments with a treatm...
print(df.isnull().sum()) #identify the null values
show id
                   0
type
                   0
title
                   0
director
                2389
cast
                 718
country
                 507
date added
                  10
                   0
release year
                   7
rating
duration
                   0
listed in
                   0
description
                   0
dtype: int64
df.dropna(inplace=True) #drop rows having missing values
df.drop duplicates(inplace=True) # remove duplicate rows
text_columns = df.select dtypes(include='object').columns
for col in text columns:
    df[col] = df[col].astype(str).str.strip().str.lower()
print(df.head())
  show id
              type
                    title
                                     director \
1
                    7:19
                           jorge michel grau
       s2
             movie
2
       s3
             movie 23:59
                                 gilbert chan
3
       s4
                        9
                                  shane acker
             movie
4
                               robert luketic
       s5
             movie
                       21
5
       s6 tv show
                       46
                                  serdar akar
                                                 cast
                                                             country \
1 demián bichir, héctor bonilla, oscar serrano, ...
                                                              mexico
2 tedd chan, stella chung, henley hii, lawrence ...
                                                           singapore
```

```
elijah wood, john c. reilly, jennifer connelly...
                                                       united states
  jim sturgess, kevin spacey, kate bosworth, aar...
                                                       united states
5 erdal beşikçioğlu, yasemin allen, melis birkan...
                                                               turkey
          date added
                      release year rating
                                            duration \
   december 23, 2016
                                              93 min
                               2016
                                    tv-ma
   december 20, 2018
                                              78 min
                               2011
                                         r
3
   november 16, 2017
                               2009
                                     pq-13
                                              80 min
     january 1, 2020
                               2008
4
                                             123 min
                                     pg-13
5
        july 1, 2017
                               2016 tv-ma 1 season
                                            listed in \
                        dramas, international movies
1
2
                 horror movies, international movies
3
   action & adventure, independent movies, sci-fi...
4
                                               dramas
5
     international tv shows, tv dramas, tv mysteries
                                          description
   after a devastating earthquake hits mexico cit...
  when an army recruit is found dead, his fellow...
  in a postapocalyptic world, rag-doll robots hi...
   a brilliant group of students become card-coun...
   a genetics professor experiments with a treatm...
if 'date added' in df.columns:
                                 #converting date into proper format
    df['date added'] = pd.to datetime(df['date added'],
errors='coerce')
    df['date added'] = df['date added'].dt.strftime('%d-%m-%Y')
<ipython-input-21-11ef0551eb15>:2: UserWarning: Could not infer
format, so each element will be parsed individually, falling back to
`dateutil`. To ensure parsing is consistent and as-expected, please
specify a format.
  df['date added'] = pd.to datetime(df['date added'], errors='coerce')
print(df.head)
<bound method NDFrame.head of</pre>
                                    show id
                                                           title
                                                type
director
          s2
                movie
                             7:19
                                    jorge michel grau
2
                             23:59
          s3
                                         gilbert chan
                movie
3
                                 9
          s4
                movie
                                          shane acker
4
                                21
                                       robert luketic
          s5
                movie
5
              tv show
                               46
                                          serdar akar
          s6
         . . .
7778
       s7779
                movie zombieland
                                      ruben fleischer
7780
       s7781
                movie
                               Z00
                                         shlok sharma
7781
       s7782
                                         peter hewitt
                movie
                              zoom
7782
       s7783
                                          josef fares
                movie
                              ZOZO
```

```
7783
      s7784
                movie
                            zubaan
                                          mozez singh
                                                     cast \
      demián bichir, héctor bonilla, oscar serrano, ...
1
2
      tedd chan, stella chung, henley hii, lawrence ...
      elijah wood, john c. reilly, jennifer connelly...
3
4
      jim sturgess, kevin spacey, kate bosworth, aar...
5
      erdal beşikçioğlu, yasemin allen, melis birkan...
      jesse eisenberg, woody harrelson, emma stone, ...
7778
      shashank arora, shweta tripathi, rahul kumar, ...
7780
      tim allen, courteney cox, chevy chase, kate ma...
7781
      imad creidi, antoinette turk, elias gergi, car...
7782
7783
      vicky kaushal, sarah-jane dias, raaghav chanan...
                                                  country
                                                           date added \
1
                                                           23-12-2016
                                                   mexico
2
                                                singapore
                                                           20-12-2018
3
                                            united states
                                                           16-11-2017
4
                                           united states
                                                           01-01-2020
5
                                                   turkey
                                                           01-07-2017
                                                           01-11-2019
7778
                                           united states
                                                           01-07-2018
7780
                                                    india
7781
                                           united states 11-01-2020
7782
      sweden, czech republic, united kingdom, denmar...
                                                           19-10-2020
                                                           02-03-2019
7783
      release year rating
                            duration \
1
                              93 min
              2016
                    tv-ma
2
              2011
                              78 min
                         r
3
              2009
                    pq - 13
                              80 min
4
              2008
                    pg-13
                             123 min
5
              2016
                    tv-ma
                            1 season
              2009
7778
                              88 min
                        r
7780
              2018
                              94 min
                    tv-ma
7781
                              88 min
              2006
                        pg
7782
              2005
                    tv-ma
                              99 min
7783
              2015
                    tv-14
                             111 min
                                               listed in \
1
                            dramas, international movies
2
                    horror movies, international movies
3
      action & adventure, independent movies, sci-fi...
4
                                                   dramas
5
        international tv shows, tv dramas, tv mysteries
. . .
7778
                                 comedies, horror movies
7780
       dramas, independent movies, international movies
```

```
7781
                     children & family movies, comedies
7782
                           dramas, international movies
7783
        dramas, international movies, music & musicals
                                            description
1
      after a devastating earthquake hits mexico cit...
2
     when an army recruit is found dead, his fellow...
3
      in a postapocalyptic world, rag-doll robots hi...
      a brilliant group of students become card-coun...
4
5
      a genetics professor experiments with a treatm...
. . .
     looking to survive in a world taken over by zo...
7778
7780
     a drug dealer starts having doubts about his t...
7781
     dragged from civilian life, a former superhero...
7782
     when lebanon's civil war deprives zozo of his ...
7783
     a scrappy but poor boy worms his way into a ty...
[4808 rows x 12 columns]>
df.columns = [col.strip().lower().replace(" ", " ") for col in
df.columns] #column
print(df.head)
title
                                           type
director
1
          s2
                movie
                             7:19
                                   jorge michel grau
2
          s3
                            23:59
                                        gilbert chan
                movie
3
          s4
                movie
                                9
                                         shane acker
4
          s5
                movie
                               21
                                      robert luketic
5
          s6
             tv show
                               46
                                         serdar akar
. . .
         . . .
                  . . .
                                     ruben fleischer
                movie zombieland
7778
       s7779
7780
       s7781
                movie
                                        shlok sharma
                              Z00
7781
       s7782
                                        peter hewitt
                movie
                             zoom
7782
       s7783
                movie
                             ZOZO
                                         josef fares
7783
       s7784
                movie
                           zubaan
                                         mozez singh
                                                   cast \
1
      demián bichir, héctor bonilla, oscar serrano, ...
2
      tedd chan, stella chung, henley hii, lawrence ...
3
      elijah wood, john c. reilly, jennifer connelly...
4
      jim sturgess, kevin spacey, kate bosworth, aar...
5
      erdal beşikçioğlu, yasemin allen, melis birkan...
      jesse eisenberg, woody harrelson, emma stone, ...
7778
7780
      shashank arora, shweta tripathi, rahul kumar, ...
7781
     tim allen, courteney cox, chevy chase, kate ma...
7782
     imad creidi, antoinette turk, elias gergi, car...
7783
     vicky kaushal, sarah-jane dias, raaghav chanan...
```

```
country date added
1
                                                            23-12-2016
                                                   mexico
2
                                                singapore
                                                           20-12-2018
3
                                            united states
                                                            16-11-2017
4
                                            united states
                                                            01-01-2020
5
                                                            01-07-2017
                                                   turkev
7778
                                                            01-11-2019
                                            united states
7780
                                                            01-07-2018
                                                     india
7781
                                            united states
                                                            11-01-2020
      sweden, czech republic, united kingdom, denmar...
                                                            19-10-2020
7782
7783
                                                     india
                                                            02-03-2019
                            duration \
      release year rating
1
              2016
                    tv-ma
                              93 min
2
              2011
                              78 min
3
              2009
                     pg-13
                              80 min
4
              2008
                     pg-13
                             123 min
5
              2016
                     tv-ma
                            1 season
. . .
               . . .
              2009
7778
                              88 min
                         r
7780
              2018
                              94 min
                     tv-ma
7781
              2006
                              88 min
                        pg
7782
              2005
                              99 min
                     tv-ma
7783
              2015
                     tv-14
                             111 min
                                                listed in \
                            dramas, international movies
1
2
                     horror movies, international movies
3
      action & adventure, independent movies, sci-fi...
4
5
        international tv shows, tv dramas, tv mysteries
7778
                                 comedies, horror movies
       dramas, independent movies, international movies
7780
7781
                      children & family movies, comedies
7782
                            dramas, international movies
7783
         dramas, international movies, music & musicals
                                              description
1
      after a devastating earthquake hits mexico cit...
2
      when an army recruit is found dead, his fellow...
3
      in a postapocalyptic world, rag-doll robots hi...
4
      a brilliant group of students become card-coun...
5
      a genetics professor experiments with a treatm...
7778
      looking to survive in a world taken over by zo...
      a drug dealer starts having doubts about his t...
7780
      dragged from civilian life, a former superhero...
7781
```

```
7782 when lebanon's civil war deprives zozo of his ...
7783 a scrappy but poor boy worms his way into a ty...

[4808 rows x 12 columns]>

if 'release_year' in df.columns:
    df['release_year'] = pd.to_numeric(df['release_year'],
    errors='coerce').astype('Int64')

df.to_csv("cleaned_netflix_data.csv", index=False) #saving it
```