PRACTICAL NO: 01 DATA WRANGLING 1

CODE:

```
import pandas as pd
df1=pd.read csv(r'E:\DSBDA\DSBDA Datasets\n movies.csv')
print(df1)
df1.head(n=5)
df1.tail(n=6)
df1.index
df1.shape
df1.columns
df1.dtypes
df1.columns.values
df1.describe(include='all')
df1.title
df1.sort_index(axis=1,ascending=False)
df1.sort_values(by='year')
df1.iloc[5]
df1[0:3]
df1.loc[:,['title','rating']]
df1.iloc[:,:3]
```

OUTPUT:

```
df1.head(n=5)
```

```
Out[45]:
          title ... votes
         Cobra Kai ... 177,031
0
         The Crown ... 199,885
1
2
     Better Call Saul ... 501,384
       Devil in Ohio ... 9,773
3
4 Cyberpunk: Edgerunners ... 15,413
[5 rows x 9 columns]
df1.tail(n=6)
Out[46]:
         title ... votes
       Breaking Bad ... 1,831,359
9951
9952
      The Imperfects ...
                           3,130
9953 The Walking Dead ... 970,067
9954
         The Crown ... 199,898
       Supernatural ... 439,601
9955
9956
       Devil in Ohio ...
                         9,786
[6 rows x 9 columns]
df1.index
Out[47]: RangeIndex(start=0, stop=9957, step=1)
```

df1.shape

```
Out[48]: (9957, 9)
df1.columns
Out[49]:
Index(['title', 'year', 'certificate', 'duration', 'genre', 'rating',
    'description', 'stars', 'votes'],
   dtype='object')
df1.dtypes
Out[50]:
title
          object
           object
year
certificate
             object
duration
             object
            object
genre
rating
           float64
description
              object
           object
stars
votes
            object
dtype: object
df1.columns.values
Out[51]:
array(['title', 'year', 'certificate', 'duration', 'genre', 'rating',
    'description', 'stars', 'votes'], dtype=object)
```

df1.describe(include='all')

Out[52]:

title year certificate ... description stars votes 6504 ... 9957 9957 8784 count 9957 9430 unique 7912 498 20 ... 9433 8615 4862 TV-MA ... Add a Plot [] top Top Gear (2020) freq 92 663 2520 ... 434 403 33 NaN ... mean NaN NaN NaN NaN NaN NaN NaN NaN ... NaN NaN NaN std min NaN NaN NaN ... NaN NaN NaN 25% NaN NaN NaN ... NaN NaN NaN 50% NaN NaN NaN ... NaN NaN NaN 75% NaN NaN NaN NaN NaN ... NaN NaN NaN NaN ... NaN NaN NaN max

[11 rows x 9 columns]

df1.title

Out[53]:

- O Cobra Kai
- 1 The Crown
- 2 Better Call Saul
- 3 Devil in Ohio
- 4 Cyberpunk: Edgerunners

9952 The Imperfects

```
The Walking Dead
9953
9954
            The Crown
9955
           Supernatural
9956
          Devil in Ohio
Name: title, Length: 9957, dtype: object
df1.sort_index(axis=1,ascending=False)
Out[54]:
      year ... certificate
0
     (2018–) ...
                  TV-14
1
    (2016–) ... TV-MA
2
   (2015–2022) ...
                    TV-MA
     (2022) ...
3
               TV-MA
4
     (2022-) ... TV-MA
     ... ...
9952 (2022-) ... TV-MA
9953 (2010–2022) ...
                      TV-MA
9954 (2016-) ... TV-MA
9955 (2005–2020) ...
                      TV-14
9956
       (2022) ... TV-MA
[9957 rows x 9 columns]
df1.sort_values(by='year')
Out[55]:
```

title ... votes

7229 The Southsiders ... 81 1445 Scarface ... 28,107

7637 Kära släkten ... 43

6499 Death at a Broadcast ... 323

6197 Midshipman Easy ... 172

...

7980 Game Changer ... NaN

7982 Detektyw Forst ... NaN

7995 Informacja zwrotna ... NaN

7999 Horton Hears a Who! ... NaN

8001 Bad Dinosaurs ... NaN

[9957 rows x 9 columns]

df1.iloc[5]

Out[56]:

title The Sandman

year (2022–)

certificate TV-MA

duration 45 min

genre Drama, Fantasy, Horror

rating 7.8

description Upon escaping after decades of imprisonment by...

stars ['Tom Sturridge, ', 'Boyd Holbrook, ', 'Patton...

votes 116,358

Name: 5, dtype: object

```
df1[0:3]
Out[57]:
       title ... votes
      Cobra Kai ... 177,031
0
      The Crown ... 199,885
1
2 Better Call Saul ... 501,384
[3 rows x 9 columns]
df1.loc[:,['title','rating']]
Out[58]:
            title rating
0
           Cobra Kai 8.5
1
           The Crown
                        8.7
2
       Better Call Saul
                        8.9
3
        Devil in Ohio
                        5.9
   Cyberpunk: Edgerunners
                              8.6
          The Imperfects
9952
         The Walking Dead
9953
                             8.1
9954
             The Crown
                          8.7
           Supernatural
9955
                          8.4
           Devil in Ohio
9956
                          5.9
```

[9957 rows x 2 columns]

```
df1.iloc[:,:3]
Out[59]:
           title
                   year certificate
0
          Cobra Kai (2018–)
                                TV-14
1
          The Crown (2016–)
                               TV-MA
      Better Call Saul (2015–2022)
2
                                    TV-MA
        Devil in Ohio
                      (2022)
3
                                TV-MA
   Cyberpunk: Edgerunners (2022–) TV-MA
         The Imperfects (2022–)
9952
                                    TV-MA
9953
        The Walking Dead (2010–2022)
9954
            The Crown (2016–)
          Supernatural (2005–2020)
9955
                                      TV-14
9956
          Devil in Ohio
                         (2022)
                                  TV-MA
[9957 rows x 3 columns]
runfile('E:/DSBDA/dsbdapr1.py', wdir='E:/DSBDA')
           title ... votes
          Cobra Kai ... 177,031
0
          The Crown ... 199,885
1
      Better Call Saul ... 501,384
2
3
        Devil in Ohio ... 9,773
   Cyberpunk: Edgerunners ... 15,413
```

...

9952 The Imperfects ... 3,130
9953 The Walking Dead ... 970,067
9954 The Crown ... 199,898
9955 Supernatural ... 439,601
9956 Devil in Ohio ... 9,786

[9957 rows x 9 columns]

OUTPUT:

```
Console I/A X
 In [38]: df1.sort_values(by='year')
title ... votes
7229 The Southsiders ... 81
1445 Scarface ... 28,107
7637 Kära släkten ... 43
6499 Death at a Broadcast ... 323
6197 Midshipman Easy ... 172
                 Game Changer ...
Detektyw Forst ...
7980
7982
                                                      NaN
          Informacja zwrotna ...
Horton Hears a Who! ...
Bad Dinosaurs ...
7995
                                                     NaN
 7999
8001
                                                     NaN
[9957 rows x 9 columns]
In [39]: df1.iloc[5]
title
                                                                               The Sandman
                                                                                    (2022- )
year
certificate
duration
                                                                                        TV-MA
                                                                                       45 min
                                                              Drama, Fantasy, Horror
genre
rating
                      Upon escaping after decades of imprisonment by...
['Tom Sturridge, ', 'Boyd Holbrook, ', 'Patton...
description
stars
Name: 5, dtype: object
In [40]: df1[0:3]
                     title untes
                                                                  IPython Console History
```

```
Ĵ
    Console 1/A X
In [36]: df1.title
                      Cobra Kai
                      The Crown
               Better Call Saul
                  Devil in Ohio
        Cyberpunk: Edgerunners
                The Imperfects
9952
               The Walking Dead
9953
9954
                     The Crown
9955
                   Supernatural
9956 Devil in Ohio
Name: title, Length: 9957, dtype: object
In [37]: df1.sort_index(axis=1,ascending=False)
             year ... certificate
         (2018- )
(2016- )
                              TV-14
                              TV-MA
      (2015-2022)
                              TV-MA
          (2022)
                              TV-MA
          (2022- )
                              TV-MA
         (2022- )
...
9952
                              TV-MA
9953
      (2010-2022)
                              TV-MA
9954
         (2016- ) ...
                              TV-MA
                              TV-14
      (2005-2020) ...
9955
           (2022) ...
9956
                              TV-MA
[9957 rows x 9 columns]
                                               IPython Console History
```

```
Console 1/A X
In [33] df1.dtypes
title
                   object
year
                   object
certificate
                  object
duration
                   object
                  object
genre
                  float64
rating
description
                  object
stars
                   object
votes
                   object
dtype: object
In [34]: df1.columns.values
array(['title', 'year', 'certificate', 'duration', 'genre', 'rating', 
'description', 'stars', 'votes'], dtype=object)
In [35]: df1.describe(include='all')
                       year certificate ... description stars
9430 6504 ... 9957 9957
498 20 ... 9433 8615
2020) TV-MA ... Add a Plot []
663 2520 ... 434 403
             title
                                                                         votes
              9957
count
                                                                          8784
              7912
unique
                                                                          4862
top
         Top Gear
                     (2020)
                                                                             30
               92
freq
                                                                             33
                                                                   NaN
               NaN
                         NaN
                                       NaN ...
                                                            NaN
mean
                                                                            NaN
                                       NaN ...
std
               NaN
                         NaN
                                                            NaN
                                                                   NaN
                                                                            NaN
               NaN
                                                            NaN
                                                                   NaN
                                                                            NaN
                         NaN
min
25%
               NaN
                         NaN
                                       NaN ...
                                                            NaN
                                                                   NaN
                                                                            NaN
                                       NaN ...
50%
               NaN
                         NaN
                                                            NaN
                                                                   NaN
                                                                            NaN
75%
               NaN
                                                                   NaN
                         NaN
                                                            NaN
                                                                            NaN
               NaN
                         NaN
                                       NaN ...
                                                            NaN
                                                                   NaN
                                                                            NaN
тах
                                                     Trython Consoler History
                                                                                                               Console 1/A X
 In [28]: df1.head(n=5)
           title ... votes
Cobra Kai ... 177,031
The Crown ... 199,885
Better Call Saul ... 501,384
Devil in Ohio ... 9,773
4 Cyberpunk: Edgerunners ... 15,413
[5 rows x 9 columns]
In [29]: df1.tail(n=6)
                     title ...
                                       votes
         Breaking Bad ... 1,831,359
The Imperfects ... 3,130
9951
9952
9953 The Walking Dead ...
                                     970,067
            The Crown ...
Supernatural ...
9954
                                     199,898
9955
                                     439,601
9956
           Devil in Ohio ...
                                       9,786
[6 rows x 9 columns]
In [30]: dfl.index
           RangeIndex(start=0, stop=9957, step=1)
 In [31]: df1.shape
     (9957, 9)
In [32]: df1.columns
 Index(['title', 'year', 'certificate', 'duration', 'genre', 'rating',
                                                     Pottun Consula History
```

```
Console 1/A X
In [43]: df1.iloc[:,:3]
                                  year certificate
                      title
                  Cobra Kai
                                (2018- )
                                              TV-14
0
1
2
3
                                 (2016-)
                                                TV-MA
                   The Crown
            Better Call Saul
                              (2015-2022)
                                                TV-MA
              Devil in Ohio
                                 (2022)
                                                TV-MA
                                 (2022-)
      Cyberpunk: Edgerunners
                                                TV-MA
9952
                                (2022- )
                                               TV-MA
             The Imperfects
9953
            The Walking Dead
                             (2010-2022)
                                               TV-MA
9954
                  The Crown
                                (2016-)
                                               TV-MA
9955
                Supernatural
                              (2005-2020)
                                                TV-14
              Devil in Ohio
                                                TV-MA
9956
                                  (2022)
[9957 rows x 3 columns]
In [44]: df1.loc[:,['title','rating']]
                      title rating
                                8.5
0
                  Cobra Kai
                  The Crown
                                 8.7
            Better Call Saul
                                8.9
             Devil in Ohio
                                 5.9
      Cyberpunk: Edgerunners
                                 8.6
9952
             The Imperfects
                                 6.3
9953
            The Walking Dead
                                8.1
9954
                 The Crown
                                8.7
9955
               Supernatural
                                 8.4
9956
              Devil in Ohio
                                 5.9
                                            Python Console History
```