**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

0 Cobra Kai ... 177,031

1 The Crown ... 199,885

2 Better Call Saul ... 501,384

3 Devil in Ohio ... 9,773

4 Cyberpunk: Edgerunners ... 15,413

[5 rows x 9 columns]

df1.tail(n=6)

Out[46]:

title ... votes

9951 Breaking Bad ... 1,831,359

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

[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

year object

certificate object

duration object

genre object

rating float64

description object

stars object

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

count 9957 9430 6504 ... 9957 9957 8784

unique 7912 498 20 ... 9433 8615 4862

top Top Gear (2020) TV-MA ... Add a Plot [] 30

freq 92 663 2520 ... 434 403 33

mean NaN NaN NaN ... NaN NaN NaN

std NaN NaN NaN ... NaN NaN NaN

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

max NaN NaN NaN ... NaN NaN NaN

[11 rows x 9 columns]

df1.title

Out[53]:

0 Cobra Kai

1 The Crown

2 Better Call Saul

3 Devil in Ohio

4 Cyberpunk: Edgerunners

9952 The Imperfects

9953 The Walking Dead

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

3 (2022) ... 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

0 Cobra Kai ... 177,031

1 The Crown ... 199,885

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

4 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

[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

2 Better Call Saul (2015–2022) TV-MA

3 Devil in Ohio (2022) TV-MA

4 Cyberpunk: Edgerunners (2022– ) TV-MA

... ... ...

9952 The Imperfects (2022– ) TV-MA

9953 The Walking Dead (2010–2022) TV-MA

9954 The Crown (2016– ) TV-MA

9955 Supernatural (2005–2020) TV-14

9956 Devil in Ohio (2022) TV-MA

[9957 rows x 3 columns]

runfile('E:/DSBDA/dsbdapr1.py', wdir='E:/DSBDA')

title ... votes

0 Cobra Kai ... 177,031

1 The Crown ... 199,885

2 Better Call Saul ... 501,384

3 Devil in Ohio ... 9,773

4 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 :**

