```
import numpy as np
import pandas as pd
df=pd.read_csv("D:/luminar/movies_cleaned_pandas.csv",header=None)
df.columns='id','name','year','rate','duration'
print(df)
```

```
          id                                                name  year

0          2                                          The Mummy  1932
\
1          3                              Orphans of the Storm  1921

2          4                              The Object of Beauty  1991

3          5                                         Night Tide  1963

4          6                               One Magic Christmas  1985

...      ...                                               ...   ...

49584  49586                                 Winter Wonderland  2013

49585  49587                Top Gear: Series 19: Africa Special  2013

49586  49588  Fireplace For Your Home: Crackling Fireplace w...  2010

49587  49589                                   Kate Plus Ei8ht  2010

49588  49590                          Kate Plus Ei8ht: Season 1  2010


           rate  duration
0      3.500000    4388.0
1      3.200000    9062.0
2      2.800000    6150.0
3      2.800000    5126.0
4      3.800000    5333.0
...         ...       ...
49584  2.800000    1812.0
49585  3.451207    6822.0
49586  3.451207    3610.0
49587  2.700000    2563.0
49588  2.700000    2563.0

[49589 rows x 5 columns]
```

1. Find row count

```
print(df.count())
```

```
id          49589
Name        49589
year        49589
rate        49589
duration    49589
dtype: int64
```

1. Remove duplicates and find row count

```
print(df.drop_duplicates().count())
```

```
id          49589
Name        49589
year        49589
rate        49589
duration    49589
dtype: int64
```

1. Sort data set by release year in des order

```
df1=df.sort_values(by='year',ascending=False)
print(df1)
```

```
          id                                                Name  year

49559  49561                                The Square (Trailer)  2014
\
43333  43335                                             Phantom  2013

45388  45390   Mako Mermaids: An H2O Adventure: Season 1: Trust  2013

45386  45388  Mako Mermaids: An H2O Adventure: Season 1: Zac...  2013

45385  45387  Mako Mermaids: An H2O Adventure: Season 1: The...  2013

...      ...                                               ...   ...

608      610                                             Cabiria  1914

20056  20058  Charlie Chaplin Collection: Shorts: A Night In...  1914

14326  14328                 Fantômas III: The Murderous Corpse  1913

42663  42665        Fantômas I: In the Shadow of the Guillotine  1913

42669  42671                     Fantômas II: Juve vs. Fantômas  1913


           rate  duration
49559  3.600000     154.0
43333  3.700000    5926.0
```

```
45388  3.451207    1458.0
45386  3.451207    1458.0
45385  3.451207    1501.0
...         ...       ...
608    2.900000    7684.0
20056  3.451207    2907.0
14326  2.600000    5432.0
42663  2.900000    3268.0
42669  2.700000    3718.0

[49589 rows x 5 columns]
```

1. Find rating mxm 5 movies name,year,rating

```
df2=df.sort_values(by='rate',ascending=False).head(5)
[['name','year','rate']]
print(df2)
```

```
                                  name  year  rate
19398      The Walking Dead: Season 2  2011   4.5
33521              Sherlock: Series 2  2012   4.5
13313                    Breaking Bad  2008   4.5
44096  Arrested Development (Trailer)  2013   4.5
36257          Breaking Bad: Season 5  2012   4.5
```

1. Find rating minimum 3 movies name,year,rtaing

```
df3=df.sort_values(by='rate').head(3) [['name','year','rate']]
print(df3)
```

```
                      name  year  rate
42159  Sun Yaar Chill Maar  2007   1.4
40933               Lagegi  2007   1.4
40825               Lagegi  2007   1.4
```

Find Each year release movie count [count desc order]

```
df4=df.groupby(by='year')['year'].count().sort_values(ascending=False)
print(df4)
```

```
year
2011    5511
2010    5107
2009    4451
2012    4339
2008    3358
        ...
1921       2
1918       1
1916       1
1915       1
2014       1
Name: year, Length: 101, dtype: int64
```

1. Each rating count [count desc order]

```
df5=df.groupby(by='rate')['rate'].count().sort_values(ascending=False)
print(df5)
```

```
rate
3.451207    38776
3.700000     1058
3.600000      967
3.800000      962
3.900000      859
3.500000      802
3.400000      698
3.300000      677
3.200000      592
4.000000      580
3.100000      545
3.000000      481
4.100000      403
2.900000      386
2.800000      316
4.200000      316
2.700000      234
2.600000      163
2.500000      156
4.300000      124
2.400000      109
2.300000       99
2.100000       74
2.200000       53
1.900000       33
4.400000       31
2.000000       24
4.500000       23
1.800000       18
1.700000       14
1.600000       10
1.400000        4
1.500000        2
Name: rate, dtype: int64
```

1. 2008 and rating above 3 [collect]

```
df6=df.loc[(df['year']==2008)&(df['rate']>3)]
print(df6)
```

```
          id                                                name  year

3187    3189                                               Elegy  2008
\
3400    3402                           Lipstick Jungle: Season 1  2008
```

```
4000    4002                          The Business of Being Born  2008

4007    4009  Pirates Who Don't Do Anything: A VeggieTales M...  2008

4015    4017                               The Tale of Despereaux  2008

...      ...                                                 ...   ...

47784  47786                   I Was Bitten: Season 1: Episode 2  2008

47793  47795                   I Was Bitten: Season 1: Episode 4  2008

47794  47796                   I Was Bitten: Season 1: Episode 5  2008

47795  47797                   I Was Bitten: Season 1: Episode 6  2008

48868  48870                                        I Was Bitten  2008


           rate  duration
3187   3.300000    6727.0
3400   3.100000    2563.0
4000   3.900000    5089.0
4007   3.700000    5123.0
4015   3.500000    5607.0
...         ...       ...
47784  3.451207    2620.0
47793  3.451207    2619.0
47794  3.451207    2618.0
47795  3.451207    2618.0
48868  3.800000    2563.0

[3254 rows x 5 columns]
```

A. row count

```
print(df6.count())
```

```
id          3254
name        3254
year        3254
rate        3254
duration    3254
dtype: int64
```

1. Find duration mxm 1 movies name,year,rating,duration

```
df7=df.sort_values(by='duration',ascending=False).head(1)
[['name','year','rate','duration']]
print(df7)
```

```
                                        name  year  rate  duration
22215  Example 8 Hour 23.976 Remote Content  2010   3.0   28813.0
```

1. Find rating mnm 1 movies name,year,rating,duration

```
df8=df.sort_values(by='rate').head(1)
[['name','year','rate','duration']]
print(df8)
```

```
                     name  year  rate  duration
42159  Sun Yaar Chill Maar  2007   1.4    2563.0
```

1. Rating above 4 and relase year above 2005

```
df9=df.loc[(df['rate']>4)&(df['year']>2005)]
print(df9)
```

```
          id                                                name  year
rate
2503    2505                                   Rang De Basanti  2006
4.1  \
2605    2607                                   Freedom Writers  2007
4.1
2656    2658                     How I Met Your Mother: Season 2  2006
4.2
2660    2662                       Prison Break: Season 2  2006
4.2
2661    2663                              Bones: Season 2  2006
4.1
...      ...                                               ...   ...
...
49547  49549                         Life With Boys: Season 1  2011
4.1
49552  49554                                      Max Steel  2013
4.1
49554  49556                   Lilyhammer: Season 1 (Recap)  2013
4.2
49569  49571                       The Short Game (Trailer)  2013
4.1
49577  49579  Transformers Prime Beast Hunters: Predacons Ri...  2013
4.2

       duration
2503     9999.0
2605     7375.0
2656     2563.0
2660     2563.0
2661     2563.0
...         ...
49547    2563.0
49552    2563.0
49554     194.0
```

```
49569     156.0
49577    3950.0

[732 rows x 5 columns]
```

A. Rating mxm movies full data

```
df10=df9.sort_values(by='rate',ascending=False)
print(df10)
```

```
          id                              name  year  rate  duration
14953  14955          Breaking Bad: Season 4  2011   4.5    2563.0
13313  13315                    Breaking Bad  2008   4.5    2563.0
12077  12079          Breaking Bad: Season 3  2010   4.5    2563.0
30424  30426                    The Avengers  2012   4.5    8575.0
44096  44098  Arrested Development (Trailer)  2013   4.5      97.0
...      ...                             ...   ...   ...       ...
35316  35318              Transformers Prime  2010   4.1    2563.0
34994  34996    Transformers Prime: Season 1  2010   4.1    2563.0
34502  34504     TEDTalks: Brave Neuro World  2011   4.1    2563.0
14336  14338                   Psych: Season 5  2010   4.1    2563.0
2503    2505                   Rang De Basanti  2006   4.1    9999.0

[732 rows x 5 columns]
```

B. Rating mnm movies full data

```
df11=df9.sort_values(by='rate')
print(df11)
```

```
          id                          name  year  rate  duration
2503    2505              Rang De Basanti  2006   4.1    9999.0
31401  31403         Last American Cowboy  2010   4.1    2563.0
31356  31358            A Good Day to Die  2010   4.1    5500.0
31143  31145  Midsomer Murders: Series 13  2010   4.1    2563.0
31087  31089        Flicka: Country Pride  2012   4.1    5529.0
...      ...                          ...   ...   ...       ...
46399  46401        The Fosters: Season 1  2013   4.5    2563.0
21909  21911             The Walking Dead  2010   4.5    2563.0
12077  12079       Breaking Bad: Season 3  2010   4.5    2563.0
14953  14955       Breaking Bad: Season 4  2011   4.5    2563.0
13313  13315                 Breaking Bad  2008   4.5    2563.0

[732 rows x 5 columns]
```

1. 2008 movies count

```
df12=df.loc[df['year']==2008].count()
print(df12)
```

```
id          3358
name        3358
```

```
year        3358
rate        3358
duration    3358
dtype: int64
```

1. 1975-2000 movies collect

```
df13=df.loc[(df['year']>1975)&(df['year']<2000)]
print(df13)
```

```
          id                                                name  year

2          4                                The Object of Beauty  1991
\
4          6                                 One Magic Christmas  1985

5          7                                    Muriel's Wedding  1994

6          8                                       Mother's Boys  1994

8         10                                        Nick of Time  1995

...      ...                                                 ...   ...

47458  47460  Goosebumps: Specials: The Haunted Mask II - Pa...  1997

47460  47462  Goosebumps: Specials: The Night of the Living ...  1997

47463  47465  Goosebumps: Specials: The Haunted Mask II - Pa...  1997

48866  48868                   The Phoenix and the Magic Carpet  1995

49580  49582                           Mumfie's White Christmas  1996


           rate  duration
2      2.800000    6150.0
4      3.800000    5333.0
5      3.500000    6323.0
6      3.400000    5733.0
8      3.400000    5333.0
...         ...       ...
47458  3.451207    1496.0
47460  3.451207    1461.0
47463  3.451207    1379.0
48866  2.500000    5337.0
49580  2.400000    1350.0

[9711 rows x 5 columns]
```

A. Row count

```
print(df13.count())
```

```
id          9711
name        9711
year        9711
rate        9711
duration    9711
dtype: int64
```

1. 1975-2000 and rating above 3.5 total row count

```
df14=df.loc[(df['year']>1975)&(df['year']<2000)&(df['rate']>3.5)].coun
t()
print(df14)
```

```
id          856
name        856
year        856
rate        856
duration    856
dtype: int64
```