```python
import numpy as np
import pandas as pd

ratings = pd.read_csv(r'C:\Users\DELL\Downloads\archive\DataCo\rating.csv')

ratings
```

```
          userId   movieId   rating              timestamp
0              1         2      3.5   2005-04-02 23:53:47
1              1        29      3.5   2005-04-02 23:31:16
2              1        32      3.5   2005-04-02 23:33:39
3              1        47      3.5   2005-04-02 23:32:07
4              1        50      3.5   2005-04-02 23:29:40
...          ...       ...      ...                   ...
20000258  138493     68954      4.5   2009-11-13 15:42:00
20000259  138493     69526      4.5   2009-12-03 18:31:48
20000260  138493     69644      3.0   2009-12-07 18:10:57
20000261  138493     70286      5.0   2009-11-13 15:42:24
20000262  138493     71619      2.5   2009-10-17 20:25:36

[20000263 rows x 4 columns]
```

```
rating
```

```
          userId   movieId   rating              timestamp
0              1         2      3.5   2005-04-02 23:53:47
1              1        29      3.5   2005-04-02 23:31:16
2              1        32      3.5   2005-04-02 23:33:39
3              1        47      3.5   2005-04-02 23:32:07
4              1        50      3.5   2005-04-02 23:29:40
...          ...       ...      ...                   ...
20000258  138493     68954      4.5   2009-11-13 15:42:00
20000259  138493     69526      4.5   2009-12-03 18:31:48
20000260  138493     69644      3.0   2009-12-07 18:10:57
20000261  138493     70286      5.0   2009-11-13 15:42:24
20000262  138493     71619      2.5   2009-10-17 20:25:36

[20000263 rows x 4 columns]
```

```python
taggings = pd.read_csv(r'C:\Users\DELL\Downloads\archive\DataCo\tag.csv')

taggings
```

```
     userId   movieId             tag              timestamp
0        18      4141     Mark Waters   2009-04-24 18:19:40
1        65       208       dark hero   2013-05-10 01:41:18
2        65       353       dark hero   2013-05-10 01:41:19
3        65       521   noir thriller   2013-05-10 01:39:43
4        65       592       dark hero   2013-05-10 01:41:18
```

```
...        ...        ...                    ...                   ...
465559  138446      55999           dragged  2013-01-23 23:29:32
465560  138446      55999   Jason Bateman    2013-01-23 23:29:38
465561  138446      55999           quirky   2013-01-23 23:29:38
465562  138446      55999              sad    2013-01-23 23:29:32
465563  138472        923    rise to power   2007-11-02 21:12:47

[465564 rows x 4 columns]
```

```python
movie = pd.read_csv(r'C:\Users\DELL\Downloads\archive\DataCo\movie.csv')
```

```python
movie
```

```
        movieId                                 title  \
0             1                     Toy Story (1995)
1             2                       Jumanji (1995)
2             3              Grumpier Old Men (1995)
3             4              Waiting to Exhale (1995)
4             5    Father of the Bride Part II (1995)
...         ...                                   ...
27273    131254          Kein Bund für's Leben (2007)
27274    131256       Feuer, Eis & Dosenbier (2002)
27275    131258                   The Pirates (2014)
27276    131260                  Rentun Ruusu (2001)
27277    131262                     Innocence (2014)

                                           genres
0      Adventure|Animation|Children|Comedy|Fantasy
1                       Adventure|Children|Fantasy
2                                   Comedy|Romance
3                             Comedy|Drama|Romance
4                                           Comedy
...                                            ...
27273                                       Comedy
27274                                       Comedy
27275                                    Adventure
27276                             (no genres listed)
27277                       Adventure|Fantasy|Horror

[27278 rows x 3 columns]
```

```python
rating.head()
```

```
   userId  movieId  rating            timestamp
0       1        2     3.5  2005-04-02 23:53:47
1       1       29     3.5  2005-04-02 23:31:16
2       1       32     3.5  2005-04-02 23:33:39
3       1       47     3.5  2005-04-02 23:32:07
4       1       50     3.5  2005-04-02 23:29:40
```

```
taggings.head()
```

```
    userId   movieId         tag              timestamp
0       18      4141    Mark Waters    2009-04-24 18:19:40
1       65       208      dark hero    2013-05-10 01:41:18
2       65       353      dark hero    2013-05-10 01:41:19
3       65       521   noir thriller   2013-05-10 01:39:43
4       65       592      dark hero    2013-05-10 01:41:18
```

```
movie.head()
```

```
    movieId                              title  \
0         1                     Toy Story (1995)
1         2                       Jumanji (1995)
2         3              Grumpier Old Men (1995)
3         4             Waiting to Exhale (1995)
4         5   Father of the Bride Part II (1995)
```

```
                                          genres
0   Adventure|Animation|Children|Comedy|Fantasy
1                    Adventure|Children|Fantasy
2                                Comedy|Romance
3                          Comedy|Drama|Romance
4                                        Comedy
```

```
movie.columns
```

```
Index(['movieId', 'title', 'genres'], dtype='object')
```

```
taggings.columns
```

```
Index(['userId', 'movieId', 'tag', 'timestamp'], dtype='object')
```

```python
del rating ['timestamp']
```

```
rating.columns
```

```
Index(['userId', 'movieId', 'rating'], dtype='object')
```

```python
row_0 = taggings.iloc[0]
type(row_0)
```

```
pandas.core.series.Series
```

```
datastructures
```

```python
row_0 = rating.iloc[0]
type(row_0)
```

```
pandas.core.series.Series
```

```python
row_0 = movie.iloc[0]
type(row_0)
```

```
pandas.core.series.Series

print(row_0)

movieId                                                    1
title                                        Toy Story (1995)
genres        Adventure|Animation|Children|Comedy|Fantasy
Name: 0, dtype: object

dataframes

taggings.head()

    userId  movieId              tag              timestamp
0       18     4141      Mark Waters  2009-04-24 18:19:40
1       65      208        dark hero  2013-05-10 01:41:18
2       65      353        dark hero  2013-05-10 01:41:19
3       65      521    noir thriller  2013-05-10 01:39:43
4       65      592        dark hero  2013-05-10 01:41:18

taggings.index

RangeIndex(start=0, stop=465564, step=1)

taggings.columns

Index(['userId', 'movieId', 'tag', 'timestamp'], dtype='object')

taggings.iloc[[0,11,500]]

      userId  movieId                  tag              timestamp
0         18     4141          Mark Waters  2009-04-24 18:19:40
11        65     1783        noir thriller  2013-05-10 01:39:43
500      342    55908    entirely dialogue  2012-01-31 18:41:16

rating.iloc[[2,67,599]]

      userId  movieId  rating
2          1       32     3.5
67         1     1997     3.5
599        7     1097     4.0

movie.iloc[[3,5,100]]

      movieId                    title                  genres
3           4   Waiting to Exhale (1995)   Comedy|Drama|Romance
5           6                 Heat (1995)   Action|Crime|Thriller
100       102          Mr. Wrong (1996)                  Comedy

desciptive statistics

rating['rating'].describe()
```

```
count    2.000026e+07
mean     3.525529e+00
std      1.051989e+00
min      5.000000e-01
25%      3.000000e+00
50%      3.500000e+00
75%      4.000000e+00
max      5.000000e+00
Name: rating, dtype: float64
```

```
rating.describe()
```

```
             userId        movieId         rating
count  2.000026e+07  2.000026e+07  2.000026e+07
mean   6.904587e+04  9.041567e+03  3.525529e+00
std    4.003863e+04  1.978948e+04  1.051989e+00
min    1.000000e+00  1.000000e+00  5.000000e-01
25%    3.439500e+04  9.020000e+02  3.000000e+00
50%    6.914100e+04  2.167000e+03  3.500000e+00
75%    1.036370e+05  4.770000e+03  4.000000e+00
max    1.384930e+05  1.312620e+05  5.000000e+00
```

```
rating.mean()
```

```
userId     69045.872583
movieId     9041.567330
rating         3.525529
dtype: float64
```

```
rating.min()
```

```
userId     1.0
movieId    1.0
rating     0.5
dtype: float64
```

```
rating.max()
```

```
userId     138493.0
movieId    131262.0
rating          5.0
dtype: float64
```

```
rating.std()
```

```
userId     40038.626653
movieId    19789.477445
rating         1.051989
dtype: float64
```

```
rating.mode()
```

```
     userId   movieId   rating
0   118205       296      4.0
```

```
rating.corr()
```

```
          userId     movieId      rating
userId   1.000000  -0.000850   0.001175
movieId -0.000850   1.000000   0.002606
rating   0.001175   0.002606   1.000000
```

```python
filterl = rating['rating']>10
print(filterl)
filterl.any()
```

```
0            False
1            False
2            False
3            False
4            False
             ...
20000258     False
20000259     False
20000260     False
20000261     False
20000262     False
Name: rating, Length: 20000263, dtype: bool
```

```
False
```

```python
filter2 = rating['rating']>0
filter2.all()
```

```
True
```

```python
movie.head()
```

```
   movieId                              title  \
0        1                   Toy Story (1995)
1        2                     Jumanji (1995)
2        3            Grumpier Old Men (1995)
3        4           Waiting to Exhale (1995)
4        5  Father of the Bride Part II (1995)


                                       genres
0  Adventure|Animation|Children|Comedy|Fantasy
1                   Adventure|Children|Fantasy
2                               Comedy|Romance
3                         Comedy|Drama|Romance
4                                       Comedy
```

```python
movie.tail()
```

```
       movieId                                title
genres
27273    131254    Kein Bund für's Leben (2007)
Comedy
27274    131256   Feuer, Eis & Dosenbier (2002)
Comedy
27275    131258                The Pirates (2014)
Adventure
27276    131260              Rentun Ruusu (2001)           (no genres
listed)
27277    131262                 Innocence (2014)   Adventure|Fantasy|
Horror
```

```
movie.describe()
```

```
            movieId
count    27278.000000
mean     59855.480570
std      44429.314697
min          1.000000
25%       6931.250000
50%      68068.000000
75%     100293.250000
max     131262.000000
```

```
movie.mode()
```

```
        movieId                                    title genres
0             1   20,000 Leagues Under the Sea (1997)   Drama
1             2                           Aladdin (1992)     NaN
2             3                           Beneath (2013)     NaN
3             4                          Blackout (2007)     NaN
4             5                          Casanova (2005)     NaN
...         ...                                      ...     ...
27273    131254                                      NaN     NaN
27274    131256                                      NaN     NaN
27275    131258                                      NaN     NaN
27276    131260                                      NaN     NaN
27277    131262                                      NaN     NaN

[27278 rows x 3 columns]
```

```
movie.min()
```

```
movieId                                                  1
title      #chicagoGirl: The Social Network Takes on a Di...
genres                                   (no genres listed)
dtype: object
```

```
movie.max()
```

```
movieId              131262
title         貞子3D (2012)
genres              Western
dtype: object
```

```
movie.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27278 entries, 0 to 27277
Data columns (total 3 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   movieId  27278 non-null  int64
 1   title    27278 non-null  object
 2   genres   27278 non-null  object
dtypes: int64(1), object(2)
memory usage: 639.5+ KB
```

```
taggings.head()
```

```
    userId  movieId              tag              timestamp
0       18     4141     Mark Waters  2009-04-24 18:19:40
1       65      208       dark hero  2013-05-10 01:41:18
2       65      353       dark hero  2013-05-10 01:41:19
3       65      521   noir thriller  2013-05-10 01:39:43
4       65      592       dark hero  2013-05-10 01:41:18
```

```
taggings.tail()
```

```
         userId  movieId              tag              timestamp
465559   138446    55999          dragged  2013-01-23 23:29:32
465560   138446    55999    Jason Bateman  2013-01-23 23:29:38
465561   138446    55999           quirky  2013-01-23 23:29:38
465562   138446    55999              sad  2013-01-23 23:29:32
465563   138472      923    rise to power  2007-11-02 21:12:47
```

data cleaning : handling missing data

```
movie.shape
```

```
(27278, 3)
```

```
movie.isnull()
```

```
       movieId  title  genres
0        False  False   False
1        False  False   False
2        False  False   False
3        False  False   False
4        False  False   False
...        ...    ...     ...
27273    False  False   False
```

```
27274      False   False    False
27275      False   False    False
27276      False   False    False
27277      False   False    False

[27278 rows x 3 columns]
```

```
movie.isnull().any().any()
```

```
False
```

```
rating.shape
```

```
(20000263, 3)
```

```
rating.isnull()
```

```
           userId   movieId   rating
0           False     False    False
1           False     False    False
2           False     False    False
3           False     False    False
4           False     False    False
...           ...       ...      ...
20000258    False     False    False
20000259    False     False    False
20000260    False     False    False
20000261    False     False    False
20000262    False     False    False

[20000263 rows x 3 columns]
```

```
rating.isnull().any().any()
```

```
False
```

```
taggings.shape
```

```
(465564, 4)
```

```
taggings.isnull().any().any()
```

```
True
```

```
taggings=taggings.dropna()
```

```
taggings
```

```
        userId   movieId              tag                 timestamp
0           18      4141      Mark Waters  2009-04-24 18:19:40
1           65       208        dark hero  2013-05-10 01:41:18
2           65       353        dark hero  2013-05-10 01:41:19
3           65       521     noir thriller  2013-05-10 01:39:43
```

```
4             65        592        dark hero    2013-05-10 01:41:18
...          ...        ...              ...                    ...
465559   138446      55999          dragged    2013-01-23 23:29:32
465560   138446      55999    Jason Bateman    2013-01-23 23:29:38
465561   138446      55999           quirky    2013-01-23 23:29:38
465562   138446      55999              sad    2013-01-23 23:29:32
465563   138472        923    rise to power    2007-11-02 21:12:47

[465548 rows x 4 columns]
```

data visualization

```
%matplotlib inline
rating.hist(column='rating',figsize=(10,5))
```

```
array([[<Axes: title={'center': 'rating'}>]], dtype=object)
```

```
rating.boxplot(column='rating',figsize=(10,5))
```

```
<Axes: title={'center': 'rating'}>
```

slicing out columns

```
rating['rating'].head()
```

```
0    3.5
1    3.5
2    3.5
3    3.5
4    3.5
Name: rating, dtype: float64
```

```
movie[['title','genres']].head()
```

```
                               title  \
0                     Toy Story (1995)
1                       Jumanji (1995)
2              Grumpier Old Men (1995)
3             Waiting to Exhale (1995)
4  Father of the Bride Part II (1995)

                                        genres
0  Adventure|Animation|Children|Comedy|Fantasy
1                   Adventure|Children|Fantasy
2                               Comedy|Romance
3                         Comedy|Drama|Romance
4                                       Comedy
```

```
rating[-10:]
```

```
          userId   movieId   rating
20000253   138493     60816      4.5
```

```
20000254    138493    61160    4.0
20000255    138493    65682    4.5
20000256    138493    66762    4.5
20000257    138493    68319    4.5
20000258    138493    68954    4.5
20000259    138493    69526    4.5
20000260    138493    69644    3.0
20000261    138493    70286    5.0
20000262    138493    71619    2.5
```