

In [1]:

```
1 import pandas as pd
2 import numpy as np
```

In []:

```
1
```

loading the data in pandas

In [2]:

```
1 df = pd.read_csv('ex1.csv')
```

In [3]:

```
1 df
```

Out[3]:

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo
3	13	14	15	16	test

In [4]:

```
1 df1 = pd.read_csv('ex1.csv',header=None) #if you don't want column header
```

In [5]:

```
1 df1
```

Out[5]:

	0	1	2	3	4
0	a	b	c	d	message
1	1	2	3	4	hello
2	5	6	7	8	world
3	9	10	11	12	foo
4	13	14	15	16	test

In [6]:

```
1 df2 = pd.read_csv('ex1.csv',names=['ALPHA','BETA','GAMMA','DELTA','MESSAGE']) #if you want to put your own col header
```

In [7]:

```
1 df2
```

Out[7]:

	ALPHA	BETA	GAMMA	DELTA	MESSAGE
0	a	b	c	d	message
1	1	2	3	4	hello
2	5	6	7	8	world
3	9	10	11	12	foo
4	13	14	15	16	test

In []:

```
1
```

In [8]:

```
1 df3 = pd.read_csv('ex1.csv',skiprows=[3])
```

In [9]:

```
1 df3
```

Out[9]:

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	13	14	15	16	test

In []:

```
1
```

handling missing values

In [10]:

```
1 df4 = pd.read_csv('ex2.csv')
```

In [11]:

```
1 df4
```

Out[11]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	world
2	three	9	10	11.0	12	foo

In [12]:

```
1 df4.isnull()
```

Out[12]:

	something	a	b	c	d	message
0	False	False	False	False	False	True
1	False	False	False	True	False	False
2	False	False	False	False	False	False

In []:

```
1
```

In [66]:

```
1 movies = pd.read_csv('movies.csv')
```

In [17]:

1	movies
1	1 2. The Godfather (1972) R 175 min Crime, Drama 9.2 100.0 The aging patriarch of an organized crime dyna... Director: Francis Ford Coppola Stars: Marlon... Votes: 1,584,782 Gross: \$134.97M
2	2 3. The Dark Knight (2008) PG-13 152 min Action, Crime, Drama 9.0 84.0 When the menace known as the Joker wreaks havo... Director: Christopher Nolan Stars: Christian... Votes: 2,260,649 Gross: \$534.86M
3	3 4. The Godfather: Part II (1974) R 202 min Crime, Drama 9.0 90.0 The early life and career of Vito Corleone in ... Director: Francis Ford Coppola Stars: Al Pac... Votes: 1,107,253 Gross: \$57.30M
4	4 5. The Lord of the Rings: The Return of the Ki... PG-13 201 min Action, Adventure, Drama 8.9 94.0 Gandalf and Aragorn lead the World of Men agai... Director: Peter Jackson Stars: Elijah Wood, ... Votes: 1,614,369 Gross: \$377.85M

In [15]:

```
1 pd.set_option('display.max_row',None)
```

In [21]:

```
1 movies.head()
```

Out[21]:

Unnamed: 0		Title	Certificate	Duration	Genre	Rate	Metascore	Description	Cast	Info
0	0	1. The Shawshank Redemption (1994)	R	142 min	Drama	9.3	80.0	Two imprisoned men bond over a number of years...	Director: Frank Darabont Stars: Tim Robbins,...	Votes: 2,295,987 Gross: \$28.34M
1	1	2. The Godfather (1972)	R	175 min	Crime, Drama	9.2	100.0	The aging patriarch of an organized crime dyna...	Director: Francis Ford Coppola Stars: Marlon...	Votes: 1,584,782 Gross: \$134.97M
2	2	3. The Dark Knight (2008)	PG-13	152 min	Action, Crime, Drama	9.0	84.0	When the menace known as the Joker wreaks havo...	Director: Christopher Nolan Stars: Christian...	Votes: 2,260,649 Gross: \$534.86M
3	3	4. The Godfather: Part II (1974)	R	202 min	Crime, Drama	9.0	90.0	The early life and career of Vito Corleone in ...	Director: Francis Ford Coppola Stars: Al Pac...	Votes: 1,107,253 Gross: \$57.30M
4	4	5. The Lord of the Rings: The Return of the Ki...	PG-13	201 min	Action, Adventure, Drama	8.9	94.0	Gandalf and Aragorn lead the World of Men agai...	Director: Peter Jackson Stars: Elijah Wood, ...	Votes: 1,614,369 Gross: \$377.85M

In [19]:

```
1 movies.tail()
```

Out[19]:

Unnamed: 0		Title	Certificate	Duration	Genre	Rate	Metascore	Description	Cast	Info
995	995	398. Scent of a Woman (1992)	R	156 min	Drama	8.0	NaN	A prep school student needing money agrees to ...	Director: Martin Brest Stars: Al Pacino, Chr...	Votes: 256,515 Gross: \$63.90M
996	996	399. Aladdin (1992)	G	90 min	Animation, Adventure, Comedy	8.0	86.0	A kindhearted street urchin and a power-hungry...	Directors: Ron Clements, John Musker Stars: ...	Votes: 367,489 Gross: \$217.35M
997	997	400. JFK (1991)	R	189 min	Drama, History, Thriller	8.0	72.0	New Orleans District Attorney Jim Garrison dis...	Director: Oliver Stone Stars: Kevin Costner,...	Votes: 139,634 Gross: \$70.41M
998	998	301. Nights of Cabiria (1957)	Not Rated	110 min	Drama	8.1	NaN	A waifish prostitute wanders the streets of Ro...	Director: Federico Fellini Stars: Giulietta ...	Votes: 42,160 Gross: \$0.75M
999	999	302. Throne of Blood (1957)	Not Rated	110 min	Drama, History	8.1	NaN	A war-hardened general, egged on by his ambiti...	Director: Akira Kurosawa Stars: Toshiro Mifu...	Votes: 45,579

In [22]:

1	movies.isnull()									
31	False	False	False	False	False	False	False	False	False	False
32	False	False	False	False	False	False	True	False	False	False
33	False	False	False	False	False	False	False	False	False	False
34	False	False	False	False	False	False	True	False	False	False
35	False	False	False	False	False	False	False	False	False	False
36	False	False	False	False	False	False	False	False	False	False
37	False	False	False	False	False	False	False	False	False	False
38	False	False	False	False	False	False	False	False	False	False
39	False	False	False	False	False	False	False	False	False	False
40	False	False	False	False	False	False	False	False	False	False
41	False	False	False	False	False	False	False	False	False	False
42	False	False	False	False	False	False	False	False	False	False
43	False	False	False	False	False	False	False	False	False	False

In []:

```
1
```

In [23]:

```
1 movies.columns
```

Out[23]:

```
Index(['Unnamed: 0', 'Title', 'Certificate', 'Duration', 'Genre', 'Rate', 'Metascore', 'Description', 'Cast', 'Info'], dtype='object')
```

In []:

```
1
```

In [24]:

```
1 movies['Rate']
```

Out[24]:

```
0      9.3
1      9.2
2      9.0
3      9.0
4      8.9
5      8.9
6      8.9
7      8.9
8      8.8
9      8.8
10     8.8
11     8.8
12     8.8
13     8.7
14     8.7
15     8.7
16     8.7
17     8.7
```

In [25]:

```
1 movies['Rate'].unique()
```

Out[25]:

```
array([9.3, 9.2, 9. , 8.9, 8.8, 8.7, 8.6, 8.5, 8.4, 8.3, 8.2, 8.1, 8. ])
```

In []:

```
1
```

In [26]:

```
1 movies['Rate'].value_counts()
```

Out[26]:

```
8.0    560
8.1    249
8.2     63
8.3     46
8.4     29
8.5     21
8.6     13
8.7      6
8.8      5
8.9      4
9.0      2
9.2      1
9.3      1
Name: Rate, dtype: int64
```

In [27]:

```
1 movies['Rate'].min()
```

Out[27]:

```
8.0
```

In [28]:

```
1 movies['Rate'].max()
```

Out[28]:

```
9.3
```

```
In [29]:
1 movies['Certificate']

Out[29]:
0          R
1          R
2        PG-13
3          R
4        PG-13
5          R
6          R
7    Approved
8        PG-13
9          R
10       PG-13
11       PG-13
12          R
13       PG-13
14       PG-13
15          R
16          R
17         PG

In [31]:
1 movies['Certificate'].unique()

Out[31]:
array(['R', 'PG-13', 'Approved', 'PG', 'Not Rated', 'G', nan, 'Passed',
       'TV-14', 'GP', 'TV-PG', 'TV-MA'], dtype=object)

In [32]:
1 movies['Certificate'].value_counts()

Out[32]:
R          345
Not Rated  200
PG-13      165
PG         105
Passed      55
Approved    45
G           41
TV-PG       7
TV-MA       7
GP           2
TV-14       1
Name: Certificate, dtype: int64

In [67]:
1 movies['Info']
66      votes: 552,500 | Gross: $11.29M
67      votes: 505,043 | Gross: $0.71M
68      votes: 1,107,298 | Gross: $25.54M
69      votes: 334,608 | Gross: $2.38M
70      votes: 304,974 | Gross: $5.32M
71      votes: 870,596 | Gross: $248.16M
72      votes: 877,721 | Gross: $44.02M
73      votes: 595,450 | Gross: $83.47M
74      votes: 774,503 | Gross: $78.90M
75      votes: 32,302
76      votes: 443,870 | Gross: $0.28M
77      votes: 106,051 | Gross: $8.18M
78      votes: 174,411
79      votes: 435,967 | Gross: $36.76M
80      votes: 197,786
81      votes: 199,210 | Gross: $0.29M
82      votes: 378,902 | Gross: $159.23M
83      votes: 41,887 | Gross: $4.19M
84      votes: 25,282
85      votes: 67,476 | Gross: $1.37M
--      --
```

In [38]:

```
1 movies.head()
```

Out[38]:

Unnamed: 0		Title	Certificate	Duration	Genre	Rate	Metascore	Description	Cast	Info
0	0	1. The Shawshank Redemption (1994)	R	142 min	Drama	9.3	80.0	Two imprisoned men bond over a number of years...	Director: Frank Darabont Stars: Tim Robbins,...	Votes: 2,295,987 Gross: \$28.34M
1	1	2. The Godfather (1972)	R	175 min	Crime, Drama	9.2	100.0	The aging patriarch of an organized crime dyna...	Director: Francis Ford Coppola Stars: Marlon...	Votes: 1,584,782 Gross: \$134.97M
2	2	3. The Dark Knight (2008)	PG-13	152 min	Action, Crime, Drama	9.0	84.0	When the menace known as the Joker wreaks havo...	Director: Christopher Nolan Stars: Christian...	Votes: 2,260,649 Gross: \$534.86M
3	3	4. The Godfather: Part II (1974)	R	202 min	Crime, Drama	9.0	90.0	The early life and career of Vito Corleone in ...	Director: Francis Ford Coppola Stars: Al Pac...	Votes: 1,107,253 Gross: \$57.30M
4	4	5. The Lord of the Rings: The Return of the Ki...	PG-13	201 min	Action, Adventure, Drama	8.9	94.0	Gandalf and Aragorn lead the World of Men agai...	Director: Peter Jackson Stars: Elijah Wood, ...	Votes: 1,614,369 Gross: \$377.85M

In [41]:

```
1 val = movies['Info'][0]
```

In [42]:

```
1 val
```

Out[42]:

'Votes: 2,295,987 | Gross: \$28.34M'

In [44]:

```
1 res = val.split('|')
```

In [45]:

```
1 res
```

Out[45]:

['Votes: 2,295,987 ', ' Gross: \$28.34M']

In [53]:

```
1 res[1]
```

Out[53]:

' Gross: \$28.34M'

In [47]:

```
1 res[1]
```

Out[47]:

' Gross: \$28.34M'

In []:

```
1
```

In [58]:

```
1 res = 'Votes: 44,141'
```

In []:

```
1 res.split()
```

In []:

```
1
```

In [68]:

```
1 def split_data(data):
2     return data.split('|')[0]
```

In [69]:

```
1 movies['Votes'] = movies['Info'].apply(split_data)
```

In [137]:

```
1 movies['Votes']
110      Votes: 228,112
111      Votes: 709,433
112      Votes: 237,865
113      Votes: 745,296
114      Votes: 592,443
115      Votes: 228,305
116      Votes: 263,491
117      Votes: 160,657
118      Votes: 294,175
119      Votes: 356,579
120      Votes: 214,849
121      Votes: 66,415
122      Votes: 143,052
123      Votes: 140,879
124      Votes: 392,283
125      Votes: 140,682
126      Votes: 157,436
127      Votes: 110,513
128      Votes: 30,676
129      Votes: 32,652
```

In [144]:

```
1 def split_gross(data):
2     if '|' in data:
3         return data.split('|')[1]
4     else:
5         return np.nan
```

In [145]:

```
1 movies['Gross'] = movies['Info'].apply(split_gross)
```

In [146]:

```
1 movies['Gross']
977      Gross: $1.48M
978      Gross: $289.92M
979      Gross: $22.46M
980      Gross: $4.18M
981      Gross: $7.22M
982      Gross: $6.20M
983      Gross: $0.78M
984      Gross: $1.65M
985      Gross: $5.60M
986      Gross: $23.16M
987      Gross: $0.00M
988      Gross: $24.48M
989      Gross: $13.42M
990      Gross: $57.14M
991      Gross: $0.52M
992      Gross: $75.08M
993      Gross: $70.91M
994      Gross: $4.50M
995      Gross: $63.90M
006      Gross: $217.35M
```

In []:

```
1
```

In [147]:

```
1 df4
```

Out[147]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	world
2	three	9	10	11.0	12	foo

In [148]:

```
1 df4.dropna()
```

Out[148]:

	something	a	b	c	d	message
2	three	9	10	11.0	12	foo

In [149]:

```
1 df4
```

Out[149]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	world
2	three	9	10	11.0	12	foo

In []:

```
1
```

In [152]:

```
1 df4.fillna("nothing")
```

Out[152]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	nothing
1	two	5	6	nothing	8	world
2	three	9	10	11.0	12	foo

In []:

```
1
```

In [153]:

```
1 data = pd.DataFrame([
2     [1.6,5.3],
3     [1,np.nan,np.nan],
4     [np.nan,np.nan,np.nan],[np.nan,6.5,3]
5 ])
```

In [154]:

```
1 data
```

Out[154]:

	0	1	2
0	1.6	5.3	NaN
1	1.0	NaN	NaN
2	NaN	NaN	NaN
3	NaN	6.5	3.0

In [156]:

```
1 data.dropna(axis=1)
```

Out[156]:

0
1
2
3

In [157]:

```
1 data.fillna(0)
```

Out[157]:

	0	1	2
0	1.6	5.3	0.0
1	1.0	0.0	0.0
2	0.0	0.0	0.0
3	0.0	6.5	3.0

In [158]:

```
1 data
```

Out[158]:

	0	1	2
0	1.6	5.3	NaN
1	1.0	NaN	NaN
2	NaN	NaN	NaN
3	NaN	6.5	3.0

In [161]:

```
1 data.fillna({0:0,1:5,2:10},inplace=True)
```

In [162]:

```
1 data
```

Out[162]:

	0	1	2
0	1.6	5.3	10.0
1	1.0	5.0	10.0
2	0.0	5.0	10.0
3	0.0	6.5	3.0

In [163]:

```
1 data
```

Out[163]:

	0	1	2
0	1.6	5.3	10.0
1	1.0	5.0	10.0
2	0.0	5.0	10.0
3	0.0	6.5	3.0

In []:

```
1
```

In [164]:

```
1 data = pd.Series([1,np.nan,3.5,np.nan,7])
```

In [165]:

```
1 data
```

Out[165]:

0	1.0
1	NaN
2	3.5
3	NaN
4	7.0

dtype: float64

In []:

```
1
```

In [167]:

```
1 data.median()
```

Out[167]:

3.5

In [168]:

```
1 data.mean()
```

Out[168]:

3.8333333333333335

In [169]:

```
1 data = pd.Series([1,np.nan,3.5,np.nan,7,10])
```

In [170]:

```
1 data.mean()
```

Out[170]:

5.375

In [171]:

```
1 data.median()
```

Out[171]:

5.25

In [172]:

```
1 data.fillna(data.mean())
```

Out[172]:

```
0    1.000
1    5.375
2    3.500
3    5.375
4    7.000
5   10.000
dtype: float64
```

In [173]:

```
1 data.fillna(data.median())
```

Out[173]:

```
0    1.00
1    5.25
2    3.50
3    5.25
4    7.00
5   10.00
dtype: float64
```

In []:

```
1
```