```
import pandas as pd
df = pd.read_csv('nyc_weather.csv')
df
```

	EST	Temperature	DewPoint	Humidity	Sea Level PressureIn
0	1/1/2016	38	23	52	30.03
1	1/2/2016	36	18	46	30.02
2	1/3/2016	40	21	47	29.86
3	1/4/2016	25	9	44	30.05
4	1/5/2016	20	-3	41	30.57
5	1/6/2016	33	4	35	30.50
6	1/7/2016	39	11	33	30.28
7	1/8/2016	39	29	64	30.20
8	1/9/2016	44	38	77	30.16
9	1/10/2016	50	46	71	29.59
10	1/11/2016	33	8	37	29.92
11	1/12/2016	35	15	53	29.85
12	1/13/2016	26	4	42	29.94
13	1/14/2016	30	12	47	29.95
14	1/15/2016	43	31	62	29.82
15	1/16/2016	47	37	70	29.52
16	1/17/2016	36	23	66	29.78
17	1/18/2016	25	6	53	29.83
18	1/19/2016	22	3	42	30.03
19	1/20/2016	32	15	49	30.13
20	1/21/2016	31	11	45	30.15
21	1/22/2016	26	6	41	30.21
22	1/23/2016	26	21	78	29.77

```
23  1/24/2016           28        11        53                  29.92

24  1/25/2016           34        18        54                  30.25

25  1/26/2016           43        29        56                  30.03

26  1/27/2016           41        22        45                  30.03

27  1/28/2016           37        20        51                  29.90

28  1/29/2016           36        21        50                  29.58

29  1/30/2016           34        16        46                  30.01

30  1/31/2016           46        28        52                  29.90


    VisibilityMiles  WindSpeedMPH PrecipitationIn  CloudCover
Events  \
0                10           8.0               0           5
NaN
1                10           7.0               0           3
NaN
2                10           8.0               0           1
NaN
3                10           9.0               0           3
NaN
4                10           5.0               0           0
NaN
5                10           4.0               0           0
NaN
6                10           2.0               0           3
NaN
7                10           4.0               0           8
NaN
8                 9           8.0               T           8
Rain
9                 4           NaN             1.8           7
Rain
10               10           NaN               0           1
NaN
11               10           6.0               T           4
NaN
12               10          10.0               0           0
NaN
13               10           5.0               T           7
NaN
14                9           5.0               T           2
NaN
15                8           7.0            0.24           7
```

```
Rain
16                  8            6.0              0.05             6  Fog-
Snow
17                  9           12.0                 T             2
Snow
18                 10           11.0                 0             1
NaN
19                 10            6.0                 0             2
NaN
20                 10            6.0                 0             1
NaN
21                  9            NaN              0.01             3
Snow
22                  1           16.0              2.31             8  Fog-
Snow
23                  8            6.0                 T             3
Snow
24                 10            3.0                 0             2
NaN
25                 10            7.0                 0             2
NaN
26                 10            7.0                 T             3
Rain
27                 10            5.0                 0             1
NaN
28                 10            8.0                 0             4
NaN
29                 10            7.0                 0             0
NaN
30                 10            5.0                 0             0
NaN

    WindDirDegrees
0              281
1              275
2              277
3              345
4              333
5              259
6              293
7               79
8               76
9              109
10             289
11             235
12             284
13             266
14             101
15             340
```

```
16              345
17              293
18              293
19              302
20              312
21               34
22               42
23              327
24              286
25              244
26              311
27              234
28              298
29              257
30              241
```

```
df['Temperature'].max()
```

```
50
```

```
df.fillna(0, inplace=True)
df['WindSpeedMPH'].mean()
```

```
6.225806451612903
```

```
import pandas as pd
weather_data = {
    'day':
['1/1/2017','1/2/2017','1/3/2017','1/4/2017','1/5/2017','1/6/2017'],
    'temperature': [32,35,28,24,32,31],
    'windspeed': [6,7,2,7,4,2],
    'event': ['Rain', 'Sunny', 'Snow','Snow','Rain', 'Sunny']
}
df = pd.DataFrame(weather_data)
df = pd.read_csv("weather_data1.csv")
df
```

```
        day  temperature  windspeed  event
0  1/1/2017           32          6   Rain
1  1/2/2017           35          7  Sunny
2  1/3/2017           28          2   Snow
3  1/4/2017           24          7   Snow
4  1/5/2017           32          4   Rain
5  1/6/2017           31          2  Sunny
```

```
df.shape # rows, columns = df.shape
```

```
(6, 4)
```

```
df.head() # df.head(3)
```

```
        day  temperature  windspeed  event
0  1/1/2017           32          6   Rain
1  1/2/2017           35          7  Sunny
2  1/3/2017           28          2   Snow
3  1/4/2017           24          7   Snow
4  1/5/2017           32          4   Rain
```

```
df.tail() # df.tail(2)
```

```
        day  temperature  windspeed  event
1  1/2/2017           35          7  Sunny
2  1/3/2017           28          2   Snow
3  1/4/2017           24          7   Snow
4  1/5/2017           32          4   Rain
5  1/6/2017           31          2  Sunny
```

```
df[1:3]
```

```
        day  temperature  windspeed  event
1  1/2/2017           35          7  Sunny
2  1/3/2017           28          2   Snow
```

Columns

```
df.columns
```

```
Index(['day', 'temperature', 'windspeed', 'event'], dtype='object')
```

```
df['day'] # or df.day
```

```
0    1/1/2017
1    1/2/2017
2    1/3/2017
3    1/4/2017
4    1/5/2017
5    1/6/2017
Name: day, dtype: object
```

```
type(df['day'])
```

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

```
df[['day','temperature']]
```

```
        day  temperature
0  1/1/2017           32
1  1/2/2017           35
2  1/3/2017           28
3  1/4/2017           24
4  1/5/2017           32
5  1/6/2017           31
```

Operations On DataFrame

```
df['temperature'].max()
```

```
35
```

```
df[df['temperature']>32]
```

```
        day  temperature  windspeed  event
1  1/2/2017           35          7  Sunny
```

```
df['day'][df['temperature'] == df['temperature'].max()] # Kinda doing SQL in pandas
```

```
1    1/2/2017
Name: day, dtype: object
```

```
df[df['temperature'] == df['temperature'].max()] # Kinda doing SQL in pandas
```

```
        day  temperature  windspeed  event
1  1/2/2017           35          7  Sunny
```

```
df[df['temperature'] == df['temperature'].max()] # Kinda doing SQL in pandas
```

```
        day  temperature  windspeed  event
1  1/2/2017           35          7  Sunny
```

```
df['event'].max() # But mean() won't work since data type is string
```

```
'Sunny'
```

```
df.describe()
```

```
       temperature  windspeed
count     6.000000   6.000000
mean     30.333333   4.666667
std       3.829708   2.338090
min      24.000000   2.000000
25%      28.750000   2.500000
50%      31.500000   5.000000
75%      32.000000   6.750000
max      35.000000   7.000000
```

set index

```
df
```

```
           temperature  windspeed  event
day
1/1/2017           32          6   Rain
```

```
1/2/2017           35          7  Sunny
1/3/2017           28          2   Snow
1/4/2017           24          7   Snow
1/5/2017           32          4   Rain
1/6/2017           31          2  Sunny
```

```
df.index
```

```
Index(['1/1/2017', '1/2/2017', '1/3/2017', '1/4/2017', '1/5/2017',
'1/6/2017'], dtype='object', name='day')
```

```
df.loc['1/2/2017']
```

```
temperature       35
windspeed          7
event          Sunny
Name: 1/2/2017, dtype: object
```

```
df.reset_index(inplace=True)
df.head()
```

```
        day  temperature  windspeed  event
0  1/1/2017           32          6   Rain
1  1/2/2017           35          7  Sunny
2  1/3/2017           28          2   Snow
3  1/4/2017           24          7   Snow
4  1/5/2017           32          4   Rain
```

```
df.set_index('event',inplace=True) # this is kind of building a hash
map using event as a key
df
```

```
            day  temperature  windspeed
event
Rain   1/1/2017           32          6
Sunny  1/2/2017           35          7
Snow   1/3/2017           28          2
Snow   1/4/2017           24          7
Rain   1/5/2017           32          4
Sunny  1/6/2017           31          2
```

```
df.loc['Snow']
```

```
            day  temperature  windspeed
event
Snow   1/3/2017           28          2
Snow   1/4/2017           24          7
```

Different Ways Of Creating Dataframe

Using csv

```
df = pd.read_csv("weather_data (2).csv")
df
```

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow

```
df=pd.read_excel("weather_data.xlsx","Sheet1")
df
```

	day	temperature	windspeed	event
0	2017-01-01	32	6	Rain
1	2017-01-02	35	7	Sunny
2	2017-01-03	28	2	Snow

Using excel

```
df=pd.read_excel("weather_data.xlsx","Sheet1")
df
```

	day	temperature	windspeed	event
0	2017-01-01	32	6	Rain
1	2017-01-02	35	7	Sunny
2	2017-01-03	28	2	Snow

Using dictionary

```
import pandas as pd
weather_data = {
    'day': ['1/1/2017','1/2/2017','1/3/2017'],
    'temperature': [32,35,28],
    'windspeed': [6,7,2],
    'event': ['Rain', 'Sunny', 'Snow']
}
df = pd.DataFrame(weather_data)
df
```

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow

Using tuples list

```
weather_data = [
    ('1/1/2017',32,6,'Rain'),
    ('1/2/2017',35,7,'Sunny'),
    ('1/3/2017',28,2,'Snow')
]
df = pd.DataFrame(data=weather_data,
columns=['day','temperature','windspeed','event'])
df
```

```
        day  temperature  windspeed  event
0  1/1/2017           32          6   Rain
1  1/2/2017           35          7  Sunny
2  1/3/2017           28          2   Snow
```

Using list of dictionaries

```
weather_data = [
    {'day': '1/1/2017', 'temperature': 32, 'windspeed': 6, 'event':
'Rain'},
    {'day': '1/2/2017', 'temperature': 35, 'windspeed': 7, 'event':
'Sunny'},
    {'day': '1/3/2017', 'temperature': 28, 'windspeed': 2, 'event':
'Snow'},

]
df = pd.DataFrame(data=weather_data,
columns=['day','temperature','windspeed','event'])
df
```

```
        day  temperature  windspeed  event
0  1/1/2017           32          6   Rain
1  1/2/2017           35          7  Sunny
2  1/3/2017           28          2   Snow
```

Read/Write CSV and Excel Files in Pandas

Read CSV

```
import pandas as pd
df = pd.read_csv("stock_data.csv")
df
```

```
  tickers            eps  revenue price         people
0   GOOGL          27.82       87   845     larry page
1     WMT           4.61      484    65           n.a.
2    MSFT             -1       85    64     bill gates
```

```
3    RIL   not available       50  1023  mukesh ambani
4    TATA            5.6       -1  n.a.     ratan tata
```

```
df = pd.read_csv("stock_data.csv", header=1) # skiprows and header are
kind of same
df
```

```
  GOOGL          27.82   87   845     larry page
0   WMT           4.61  484    65           n.a.
1  MSFT             -1   85    64     bill gates
2  RIL   not available   50  1023  mukesh ambani
3  TATA            5.6   -1  n.a.     ratan tata
```

```
df = pd.read_csv("stock_data.csv", header=None, names =
["ticker","eps","revenue","people"])
df
```

```
                ticker      eps revenue         people
tickers            eps  revenue   price         people
GOOGL            27.82       87     845     larry page
WMT               4.61      484      65           n.a.
MSFT                -1       85      64     bill gates
RIL      not available       50    1023  mukesh ambani
TATA              5.6        -1    n.a.     ratan tata
```

```
df = pd.read_csv("stock_data.csv",  nrows=2)
df
```

```
  tickers    eps  revenue  price      people
0   GOOGL  27.82       87    845  larry page
1     WMT   4.61      484     65        n.a.
```

```
df = pd.read_csv("stock_data.csv", na_values=["n.a.", "not
available"])
df
```

```
  tickers    eps  revenue   price         people
0   GOOGL  27.82       87   845.0     larry page
1     WMT   4.61      484    65.0            NaN
2    MSFT  -1.00       85    64.0     bill gates
3    RIL     NaN       50  1023.0  mukesh ambani
4    TATA   5.60       -1     NaN     ratan tata
```

```
df = pd.read_csv("stock_data.csv",  na_values={
        'eps': ['not available'],
        'revenue': [-1],
        'people': ['not available','n.a.']
    })
df
```

```
  tickers    eps  revenue price         people
0   GOOGL  27.82     87.0   845     larry page
1     WMT   4.61    484.0    65            NaN
2    MSFT  -1.00     85.0    64     bill gates
3     RIL    NaN     50.0  1023  mukesh ambani
4    TATA   5.60      NaN  n.a.     ratan tata
```

Write to CSV

```
df.to_csv("new.csv", index=False)

df.columns

Index(['tickers', 'eps', 'revenue', 'price', 'people'],
dtype='object')

df.to_csv("new.csv",header=False)

df.to_csv("new.csv", columns=["tickers","price"], index=False)
```

Read Excel

```
df = pd.read_excel("stock_data.xlsx","Sheet1")
df
```

```
  tickers            eps  revenue price         people
0   GOOGL          27.82       87   845     larry page
1     WMT           4.61      484    65           n.a.
2    MSFT             -1       85    64     bill gates
3     RIL  not available       50  1023  mukesh ambani
4    TATA            5.6       -1  n.a.     ratan tata
```

```
def convert_people_cell(cell):
    if cell=="n.a.":
        return 'Sam Walton'
    return cell

def convert_price_cell(cell):
    if cell=="n.a.":
        return 50
    return cell
df = pd.read_excel("stock_data.xlsx","Sheet1", converters= {
        'people': convert_people_cell,
        'price': convert_price_cell
    })
df
```

```
  tickers            eps  revenue  price         people
0   GOOGL          27.82       87    845     larry page
1     WMT           4.61      484     65     Sam Walton
2    MSFT             -1       85     64     bill gates
```

```
3    RIL   not available       50   1023  mukesh ambani
4    TATA            5.6       -1     50     ratan tata
```