Lecture17

April 4, 2024

[2]:
```
# !pip install pandas
```

[4]:
```
import pandas as pd
```

[5]:
```
# create pd.Series from list or dict
# create pd.DataFrame from nested list or from dict
# read data as dataframe (pd.read_csv
# dataframe attributes (df.index, df.columns, df.values)
# prepare dataframe (set_index, reset_index, df.columns.str.replace)
```

0.1 Combine Data

0.1.1 Concat

[7]:
```
import numpy as np
df1=pd.DataFrame(np.arange(9).reshape(3,3), columns=['a','b','c'],␣
 ↪index=['one','two','three'])
df1
```

[7]:

	a	b	c
one	0	1	2
two	3	4	5
three	6	7	8

[9]:
```
df2=pd.DataFrame(np.arange(6).reshape(3,2), columns=['d','e'],␣
 ↪index=['three','two','One'])
df2
```

[9]:

	d	e
three	0	1
two	2	3
One	4	5

[22]:
```
pd.concat([df1, df2], axis=1, join='outer')
```

[22]:

	a	b	c	d	e
one	0.0	1.0	2.0	NaN	NaN
two	3.0	4.0	5.0	2.0	3.0

```
three  6.0  7.0  8.0  0.0  1.0
One    NaN  NaN  NaN  4.0  5.0
```

[15]:
```
pd.concat([df1, df2], axis=0, join='inner')
```

[15]:
```
Empty DataFrame
Columns: []
Index: [one, two, three, three, two, One]
```

0.1.2 Merge

[17]:
```
df3=pd.DataFrame([['a','b','c'],
                  ['d','e','f'],
                  ['g','h','i']], columns=['col1','col2','col3'])
df3
```

[17]:

	col1	col2	col3
0	a	b	c
1	d	e	f
2	g	h	i

[19]:
```
df4=pd.DataFrame([['x',1,'i'],
                  ['e',2,'f'],
                  ['b',3,'e'],
                  ['z',4,'h']], columns=['col2','col4','col5'])
df4
```

[19]:

	col2	col4	col5
0	x	1	i
1	e	2	f
2	b	3	e
3	z	4	h

[26]:
```
pd.merge(df3,df4, on='col2', how='inner')
```

[26]:

	col1	col2	col3	col4	col5
0	a	b	c	3	e
1	d	e	f	2	f

[27]:
```
pd.merge(df3,df4, on='col2', how='outer')
```

[27]:

	col1	col2	col3	col4	col5
0	a	b	c	3.0	e
1	d	e	f	2.0	f
2	g	h	i	NaN	NaN
3	NaN	x	NaN	1.0	i
4	NaN	z	NaN	4.0	h

[28]:
```
pd.merge(df3,df4, on='col2', how='right')
```

[28]:

	col1	col2	col3	col4	col5
0	NaN	x	NaN	1	i
1	d	e	f	2	f
2	a	b	c	3	e
3	NaN	z	NaN	4	h

[29]:
```
pd.merge(df3,df4, on='col2', how='left')
```

[29]:

	col1	col2	col3	col4	col5
0	a	b	c	3.0	e
1	d	e	f	2.0	f
2	g	h	i	NaN	NaN

[30]:
```
pd.merge(df3,df4, left_on='col2', right_on='col5', how='outer')
```

[30]:

	col1	col2_x	col3	col2_y	col4	col5
0	a	b	c	NaN	NaN	NaN
1	d	e	f	b	3.0	e
2	NaN	NaN	NaN	e	2.0	f
3	g	h	i	z	4.0	h
4	NaN	NaN	NaN	x	1.0	i

0.2 Data Selection

[ ]:
```
# df.colname[rowname]
# df[colname][rowname]
# df.loc[rowname, colname]
# df.iloc[rowid, colid]
```

[31]:
```
df1
```

[31]:

	a	b	c
one	0	1	2
two	3	4	5
three	6	7	8

[36]:
```
df1.a['two']
```

[36]: 3

[38]:
```
df1['a']['two']
```

[38]: 3

[40]:
```
df1[['a','c']]
```

[40]:

	a	c
one	0	2
two	3	5
three	6	8

[42]:
```
df1[df1.columns[-2:]]
```

[42]:

	b	c
one	1	2
two	4	5
three	7	8

[46]:
```
df1
```

[46]:

	a	b	c
one	0	1	2
two	3	4	5
three	6	7	8

[45]:
```
# df.loc[rowname, colname]
df1.loc['two','a']
```

[45]: 3

[48]:
```
# df1.loc[['one','three'], 'a']
df1.loc[['one','three'], ['a','c']]
```

[48]:

	a	c
one	0	2
three	6	8

[49]:
```
df1.loc[['one','three'], :]
```

[49]:

	a	b	c
one	0	1	2
three	6	7	8

[50]:
```
df1.loc[:,:]
```

[50]:

	a	b	c
one	0	1	2
two	3	4	5
three	6	7	8

[51]:
```
df1.loc[:, 'a']
```

[51]:
```
one      0
two      3
three    6
Name: a, dtype: int64
```

[54]:
```
# df1.loc[df1.b>2,:]
df1.loc[(df1.b>2)&(df1.c>5),:]
```

[54]:
```
       a  b  c
three  6  7  8
```

[55]:
```
# df.iloc[rowid, colid]
df1.iloc[:,:]
```

[55]:
```
       a  b  c
one    0  1  2
two    3  4  5
three  6  7  8
```

[58]:
```
# df1.iloc[:2,:]
df1.iloc[:,-1]
```

[58]:
```
one      2
two      5
three    8
Name: c, dtype: int64
```

[ ]:

[ ]: