Pandas Cheat Sheet: Reshaping data

Date	Open	High	Low	Close	Volume	Dividends	Stock Spli	ticker
01/02/2018	65.43502	65.89118	65.35737117	65.60972	1047800	0	0	A
01/02/2018	53.9422	55.09967	53.79252158	55.04977	2928900	0	0	AA
01/02/2018	60.22431	61.49961	60.18624689	61.25216	436200	0	0	B
01/02/2018	281.8477	283.0294	281.5141014	282.8864	2978900	0	0	BA
01/02/2018	176.399	184.1	175.6999969	183.65	29916900	0	0	BABA
01/03/2018	65.62914	67.44408	65.60972223	67.27908	1698900	0	0	A
01/03/2018	54.80032	55.02982	52.84459199	54.38124	4100000	0	0	AA
01/03/2018	61.06182	61.52815	60.65257804	60.91906	194800	0	0	B
01/03/2018	282.0287	284.4684	281.5808089	283.8013	3211200	0	0	BA
01/03/2018	185.19	185.635	181.3999939	184	20121900	0	0	BABA
01/04/2018	67.49259	67.76435	66.75496824	66.77438	2230700	0	0	A
01/04/2018	54.69056	55.30921	53.95716487	54.5808	3555100	0	0	AA
01/04/2018	61.34733	62.18483	61.34732986	61.95642	149100	0	0	B
01/04/2018	283.9347	284.3922	281.5808231	282.7244	4171700	0	0	BA
01/04/2018	185.9	187.747	184.4299927	185.71	19473800	0	0	BABA

pivot

Date	Close	ticker
01/02/2018	65.60972	A
01/02/2018	55.04977	AA
01/02/2018	61.25216	B
01/02/2018	282.8864	BA
01/02/2018	183.65	BABA
01/03/2018	67.27908	A
01/03/2018	54.38124	AA
01/03/2018	60.91906	B
01/03/2018	283.8013	BA
01/03/2018	184	BABA
01/04/2018	66.77438	A
01/04/2018	54.5808	AA
01/04/2018	61.95642	B
01/04/2018	282.7244	BA
01/04/2018	185.71	BABA

```
df.pivot(index='Date', columns='ticker', values=['Close'])
```

	Close				
ticker	A	AA	B	BA	BABA
Date					
01/02/2018	65.60972	55.04977	61.25216	282.8864	183.65
01/03/2018	67.27908	54.38124	60.91906	283.8013	184
01/04/2018	66.77438	54.5808	61.95642	282.7244	185.71

unstack

	A	AA	B	BA	BABA
Date					
01/02/2018	65.60972	55.04977	61.25216	282.8864	183.65
01/03/2018	67.27908	54.38124	60.91906	283.8013	184
01/04/2018	66.77438	54.5808	61.95642	282.7244	185.71

```
pivot_df.unstack().reset_index()
```

level_0	Date	0
A	01/02/2018	65.60972
A	01/03/2018	67.27908
A	01/04/2018	66.77438
AA	01/02/2018	55.04977
AA	01/03/2018	54.38124
AA	01/04/2018	54.5808
B	01/02/2018	61.25216
B	01/03/2018	60.91906
B	01/04/2018	61.95642
BA	01/02/2018	282.8864
BA	01/03/2018	283.8013
BA	01/04/2018	282.7244
BABA	01/02/2018	183.65
BABA	01/03/2018	184
BABA	01/04/2018	185.71

melt

Date	A	AA	B	BA	BABA
01/02/2018	65.60972	55.04977	61.25216	282.8864	183.65
01/03/2018	67.27908	54.38124	60.91906	283.8013	184
01/04/2018	66.77438	54.5808	61.95642	282.7244	185.71

```
pd.melt(pivot_df,
        id_vars=['Date'],
        value_vars=['A', 'AA', 'B', 'BA', 'BABA'],
        var_name = 'ticker',
        value_name='Close'
        )
```

Date	ticker	Close
01/02/2018	A	65.60972
01/03/2018	A	67.27908
01/04/2018	A	66.77438
01/02/2018	AA	55.04977
01/03/2018	AA	54.38124
01/04/2018	AA	54.5808
01/02/2018	B	61.25216
01/03/2018	B	60.91906
01/04/2018	B	61.95642
01/02/2018	BA	282.8864
01/03/2018	BA	283.8013
01/04/2018	BA	282.7244
01/02/2018	BABA	183.65
01/03/2018	BABA	184
01/04/2018	BABA	185.71

```
pd.melt(pivot_df,
        id_vars=['Date'],
        value_vars=['A', 'AA', 'B', 'BA', 'BABA'],
        var_name = 'ticker',
        value_name='Close'
        ).sort_values(by='Date')
```

Date	ticker	Close
01/02/2018	A	65.60972
01/02/2018	AA	55.04977
01/02/2018	B	61.25216
01/02/2018	BA	282.8864
01/02/2018	BABA	183.65
01/03/2018	A	67.27908
01/03/2018	AA	54.38124
01/03/2018	B	60.91906
01/03/2018	BA	283.8013
01/03/2018	BABA	184
01/04/2018	A	66.77438
01/04/2018	AA	54.5808
01/04/2018	B	61.95642
01/04/2018	BA	282.7244
01/04/2018	BABA	185.71

stack

```
df_for_stack[ ['Open' ,'Close']].stack()
```

	Open	Close
Date		
01/02/2018	65.43502	65.60972
01/03/2018	65.62914	67.27908
01/04/2018	67.49259	66.77438

Date		
01/02/2018	Open	65.43502
	Close	65.60972
01/03/2018	Open	65.62914
	Close	67.27908
01/04/2018	Open	67.49259
	Close	66.77438

Pandas Cheat Sheet: Reshaping data

Date	Open	High	Low	Close	Volume	Dividends	Stock Spli	ticker
01/02/2018	65.43502	65.89118	65.35737117	65.60972	1047800	0	0	A
01/02/2018	53.9422	55.09967	53.79252158	55.04977	2928900	0	0	AA
01/02/2018	60.22431	61.49961	60.18624689	61.25216	436200	0	0	B
01/02/2018	281.8477	283.0294	281.5141014	282.8864	2978900	0	0	BA
01/02/2018	176.399	184.1	175.6999969	183.65	29916900	0	0	BABA
01/03/2018	65.62914	67.44408	65.60972223	67.27908	1698900	0	0	A
01/03/2018	54.80032	55.02982	52.84459199	54.38124	4100000	0	0	AA
01/03/2018	61.06182	61.52815	60.65257804	60.91906	194800	0	0	B
01/03/2018	282.0287	284.4684	281.5808089	283.8013	3211200	0	0	BA
01/03/2018	185.19	185.635	181.3999939	184	20121900	0	0	BABA
01/04/2018	67.49259	67.76435	66.75496824	66.77438	2230700	0	0	A
01/04/2018	54.69056	55.30921	53.95716487	54.5808	3555100	0	0	AA
01/04/2018	61.34733	62.18483	61.34732986	61.95642	149100	0	0	B
01/04/2018	283.9347	284.3922	281.5808231	282.7244	4171700	0	0	BA
01/04/2018	185.9	187.747	184.4299927	185.71	19473800	0	0	BABA

pivot

Date	Close	ticker
01/02/2018	65.60972	A
01/02/2018	55.04977	AA
01/02/2018	61.25216	B
01/02/2018	282.8864	BA
01/02/2018	183.65	BABA
01/03/2018	67.27908	A
01/03/2018	54.38124	AA
01/03/2018	60.91906	B
01/03/2018	283.8013	BA
01/03/2018	184	BABA
01/04/2018	66.77438	A
01/04/2018	54.5808	AA
01/04/2018	61.95642	B
01/04/2018	282.7244	BA
01/04/2018	185.71	BABA

```
df.pivot(index='Date', columns='ticker', values=['Close'])
```

	Close				
ticker	A	AA	B	BA	BABA
Date					
01/02/2018	65.60972	55.04977	61.25216	282.8864	183.65
01/03/2018	67.27908	54.38124	60.91906	283.8013	184
01/04/2018	66.77438	54.5808	61.95642	282.7244	185.71

pivot_table

Date	Close	ticker
01/02/2018	65.60972	A
01/02/2018	55.04977	AA
01/02/2018	61.25216	B
01/02/2018	282.8864	BA
01/02/2018	183.65	BABA
01/03/2018	67.27908	A
01/03/2018	54.38124	AA
01/03/2018	60.91906	B
01/03/2018	283.8013	BA
01/03/2018	184	BABA
01/04/2018	66.77438	A
01/04/2018	54.5808	AA
01/04/2018	61.95642	B
01/04/2018	282.7244	BA
01/04/2018	185.71	BABA

```
pd.pivot_table(df,
               index='Date',
               columns='ticker',
               values=['Close'],
               aggfunc={'Close':[np.max]}
              )
```

	Close				
	amax				
ticker	A	AA	B	BA	BABA
Date					
01/02/2018	65.60971832	55.04977	61.25216	282.8864	183.65
01/03/2018	67.27908325	54.38124	60.91906	283.8013	184
01/04/2018	66.77438354	54.5808	61.95642	282.7244	185.71

```
pd.pivot_table(df,
               index='Date',
               values=['Close', 'High'],
               aggfunc={'Close':[np.mean, np.std],
                        'High':[np.min]
                       }
              )
```

	Close		High
	mean	std	amin
Date			
01/02/2018	129.6896	100.923245	55.09967
01/03/2018	130.0761	101.236219	55.02982
01/04/2018	130.3492	100.924734	55.30921