

8.1 - Groupby

July 12, 2017

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: df = pd.concat([pd.read_csv('quotes2.csv')]*100)
```

```
In [3]: df.head(20)
```

```
Out[3]:
```

		time	ticker	bid	ask
0	20160525	13:30:00.023	GOOG	720.50	720.93
1	20160525	13:30:00.023	MSFT	51.95	51.95
2	20160525	13:30:00.041	MSFT	51.95	51.95
3	20160525	13:30:00.048	GOOG	720.50	720.93
4	20160525	13:30:00.048	GOOG	720.50	720.93
5	20160525	13:30:00.048	GOOG	720.50	720.93
6	20160525	13:30:00.048	GOOG	720.50	720.93
7	20160525	13:30:00.072	GOOG	720.50	720.88
8	20160525	13:30:00.075	AAPL	98.55	98.56
9	20160525	13:30:00.076	AAPL	98.55	98.56
10	20160525	13:30:00.076	AAPL	98.55	98.56
11	20160525	13:30:00.076	AAPL	98.55	98.56
12	20160525	13:30:00.078	MSFT	51.95	51.95
13	20160525	13:30:00.078	MSFT	51.95	51.95
14	20160525	13:30:00.078	MSFT	51.95	51.95
15	20160525	13:30:00.078	MSFT	51.92	51.95
16	20160525	13:30:00.079	MSFT	51.92	51.95
17	20160525	13:30:00.080	AAPL	98.55	98.56
18	20160525	13:30:00.084	AAPL	98.55	98.56
19	20160525	13:30:00.086	AAPL	98.55	98.63

```
In [4]: %timeit df.groupby('ticker').apply(np.mean)
```

10 loops, best of 3: 91.2 ms per loop

```
In [5]: df.groupby('ticker').apply(np.mean)
```

```
Out[5]:
```

	bid	ask
ticker		

AAPL	98.60525	98.617500
GOOG	720.50000	720.921667
MSFT	51.93700	51.950000

```
In [7]: %timeit df.groupby('ticker').aggregate(np.mean)
```

1000 loops, best of 3: 1.19 ms per loop

```
In [8]: %timeit df.groupby('ticker').transform(lambda x : 2 * x)
```

100 loops, best of 3: 8.48 ms per loop