

Mar-3



```
%pyspark
import numpy as np
import pandas as pd
```

FINISHED ▶ ⌵ ⌶ ⚙

Took 0 sec. Last updated by anonymous at March 09 2017, 7:07:35 PM.

```
%pyspark
df = pd.DataFrame({'key1': ['a','a','b','b','a'], 'key2' : ['one','two','one','two','one'], 'data1' : np.random.
```

FINISHED ▶ ⌕ 📖 ⚙️

Took 0 sec. Last updated by anonymous at March 09 2017, 7:07:39 PM.

```
%pyspark
df
```

FINISHED ▶ ⌵ 📖 ⚙️

	data1	data2	key1	key2
0	1.190083	2.016268	a	one
1	1.046512	1.580925	a	two
2	0.261075	-0.508882	b	one
3	-1.517274	1.170868	b	two
4	-1.096952	0.393116	a	one

Took 0 sec. Last updated by anonymous at March 09 2017, 7:07:55 PM.

```
%pyspark
grouped = df['data1'].groupby(df['key1'])
```

FINISHED ▶ ⌵ 📖 ⚙️

Took 0 sec. Last updated by anonymous at March 09 2017, 7:11:29 PM.

```
%pyspark
grouped
```

FINISHED ▶ ⌘ 📖 ⚙️

```
<pandas.core.groupby.SeriesGroupBy object at 0x7ffb5c811510>
```


Took 0 sec. Last updated by anonymous at March 09 2017, 7:11:48 PM.

```
%pyspark
grouped.mean()
```

FINISHED ▶ ⌵ ⌶ ⚙

```
key1
a    0.379881
b   -0.628099
Name: data1, dtype: float64
```


Took 0 sec. Last updated by anonymous at March 09 2017, 7:12:16 PM.

```
%pyspark
means = df['data1'].groupby([df['key1'],df['key2']]).mean()
```

FINISHED ▶ 🔍 📖 ⚙️


```
%pyspark
df.groupby(['key1', 'key2']).size()
```

default ▼

Took 0 sec. Last updated by anonymous at March 09 2017, 7:22:14 PM.

Took 0 sec. Last updated by anonymous at March 09 2017, 7:23:23 PM.

Took 0 sec. Last updated by anonymous at March 09 2017, 7:24:29 PM.

3/4

Took 0 sec. Last updated by anonymous at March 09 2017, 7:26:00 PM.

Zeppelin

Mar 3

```
%pyspark
df.dtypes
data1      float64
data2      float64
key1       object
key2       object
dtype: object
```

Took 0 sec. Last updated by anonymous at March 09 2017, 7:26:31 PM.

FINISHED





default ▾

```
%pyspark
grouped = df.groupby(df.dtypes, axis=1)
dict(list(grouped))
```

```
{dtype('O'):   key1 key2
0    a  one
1    a  two
2    b  one
3    b  two
4    a  one, dtype('float64'):   data1   data2
0  1.190083  2.016268
1  1.046512  1.580925
2  0.261075 -0.508882
3 -1.517274  1.170868
4 -1.096952  0.393116}
```

Took 0 sec. Last updated by anonymous at March 09 2017, 7:27:35 PM.

FINISHED





READY

