DS-670 Group Ope...

```
%pyspark
from pandas import Series, DataFrame
import pandas as pd
import numpy as np
```

FINISHED

```
%pyspark
s = Series(np.random.randn(6))
s[::2] = np.nan
s
```

FINISHED

```
0         NaN
1    0.783313
2         NaN
3    1.175026
4         NaN
5    0.579615
dtype: float64
```

```
%pyspark
states = ['Ohio', 'New York', 'Vermont', 'Florida', 'Oregon', 'Nevada', 'Califronia', 'Idal
group_key = ['East'] * 4 + ['West'] * 4
data = Series(np.random.randn(8), index=states)
data[['Vermont','Nevada','Idaho']] = np.nan
data
```

FINISHED

```
Ohio          0.120148
New York      0.595034
Vermont            NaN
Florida      -1.983958
Oregon       -1.399858
Nevada             NaN
Califronia    0.160380
Idaho              NaN
dtype: float64
```

```
%pyspark
data.groupby(group_key).mean()
```

FINISHED

```
East   -0.422925
West   -0.619739
dtype: float64
```

```
%pyspark
fill_mean = lambda g : g.fillna(g.mean())
data.groupby(group_key).apply(fill_mean)
```

FINISHED

```
Ohio          0.120148
New York      0.595034
Vermont      -0.422925
Florida      -1.983958
Oregon       -1.399858
Nevada       -0.619739
Califronia    0.160380
Idaho        -0.619739
dtype: float64
```

```
%pyspark

fill_values = {'East': 0.5, 'West': -1}
fill_func = lambda g: g.fillna(fill_values[g.name])
data.groupby(group_key).apply(fill_func)
```

FINISHED

```
Ohio          0.120148
New York      0.595034
Vermont       0.500000
Florida      -1.983958
Oregon       -1.399858
Nevada       -1.000000
Califronia    0.160380
Idaho        -1.000000
dtype: float64
```

```
%pyspark
```

FINISHED