Python for Data Science

Missing Data

PIERIAN DATA
www.pieriandata.com

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

Pandas - Missing Data

In [84]:
```
import numpy as np
import pandas as pd
```

In [114]:
```
d = {'A':[1,2,np.nan],'B':[5,np.nan,np.nan],'C':[1,2,3]}
```

In [115]:
```
df = pd.DataFrame(d)
```

In [116]:
```
df
```

Out[116]:

	A	B	C
0	1.0	5.0	1
1	2.0	NaN	2
2	NaN	NaN	3

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]

In [116]:
```
df
```

Out[116]:

	A	B	C
0	1.0	5.0	1
1	2.0	NaN	2
2	NaN	NaN	3

In [117]:
```
df.dropna()
```

Out[117]:

```
Signature: df.dropna(axis=0, how='any', thresh=None, subset=None, inplace=False)
Docstring:
Return object with labels on given axis omitted where alternately any
or all of the data are missing
```

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]

In [116]:
```
df
```

Out[116]:

	A	B	C
0	1.0	5.0	1
1	2.0	NaN	2
2	NaN	NaN	3

In [117]:
```
df.dropna()
```

Out[117]:

	A	B	C
0	1.0	5.0	1

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]

In [116]:
```
df
```

Out[116]:

	A	B	C
0	1.0	5.0	1
1	2.0	NaN	2
2	NaN	NaN	3

In [118]:
```
df.dropna(axis=1)
```

Out[118]:

	C
0	1
1	2
2	3

In [ ]:

2	NaN	NaN	3

In [118]:
```
df.dropna(axis=1)
```

Out[118]:

	C
0	1
1	2
2	3

In [119]:
```
df.dropna()
```

Out[119]:

	A	B	C
0	1.0	5.0	1

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]

In [119]:
```
df.dropna()
```

Out[119]:

	A	B	C
0	1.0	5.0	1

In [120]:
```
df.dropna
```

Out[120]:

```
Signature: df.dropna(axis=0, how='any', thresh=None, subset=None, inplace=False)
Docstring:
Return object with labels on given axis omitted where alternately any
or all of the data are missing
```

2	NaN	NaN	3

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]

In [119]:
```
df.dropna()
```

Out[119]:

	A	B	C
0	1.0	5.0	1

In [120]:
```
df.dropna
```

Out[120]:
```
axis : {0 or 'index', 1 or 'columns'}, or tuple/list thereof
    Pass tuple or list to drop on multiple axes
how : {'any', 'all'}
    * any : if any NA values are present, drop that label
    * all : if all values are NA, drop that label
thresh : int, default None
    int value : require that many non-NA values
subset : array-like
    Labels along other axis to consider, e.g. if you are dropping rows
    these would be a list of columns to include
inplace : boolean, default False
```

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

1	2
2	3

In [119]:
```
df.dropna()
```

Out[119]:

	A	B	C
0	1.0	5.0	1

In [121]:
```
df.dropna()
```

Out[121]:

	A	B	C
0	1.0	5.0	1

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help Python [conda env:py35]

In [119]:
```
df.dropna()
```

Out[119]:

	A	B	C
0	1.0	5.0	1

In [122]:
```
df.dropna(thresh=2)
```

Out[122]:

	A	B	C
0	1.0	5.0	1
1	2.0	NaN	2

thresh keeps no. of "non-NaN" values, means those containing at least 2 "non-NaN" values.

In [ ]:

1	2.0	NaN	2

In [123]:
```
df
```

Out[123]:

	A	B	C
0	1.0	5.0	1
1	2.0	NaN	2
2	NaN	NaN	3

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]

1	2.0	NaN	2

In [123]:
```
df.fillna()
```

Out[123]:

```
Signature: df.fillna(value=None, method=None, axis=None, inplace=False, limit=Non
e, downcast=None, **kwargs)
Docstring:
Fill NA/NaN values using the specified method
```

2	NaN	NaN	3

In [ ]:

1	2.0	NaN	2

In [124]:
```
df.fillna(value='FILL VALUE')
```

Out[124]:

	A	B	C
0	1	5	1
1	2	FILL VALUE	2
2	FILL VALUE	FILL VALUE	3

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]

In [124]:
```
df.fillna(value='FILL VALUE')
```

Out[124]:

	A	B	C
0	1	5	1
1	2	FILL VALUE	2
2	FILL VALUE	FILL VALUE	3

In [125]:
```
df['A']
```

Out[125]:
```
0    1.0
1    2.0
2    NaN
Name: A, dtype: float64
```

In [ ]:

Pandas

localhost:8888/notebooks/Pandas.ipynb

File Edit View Insert Cell Kernel Widgets Help | Python [conda env:py35]

In [124]:
```
df.fillna(value='FILL VALUE')
```

Out[124]:

	A	B	C
0	1	5	1
1	2	FILL VALUE	2
2	FILL VALUE	FILL VALUE	3

In [126]:
```
df['A'].fillna(value=df['A'].mean())
```

Out[126]:
```
0    1.0
1    2.0
2    1.5
Name: A, dtype: float64
```

In [ ]: