

Module 02

Missing Data

Data Science Developer

Create DataFrame with NaN Values

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

```
In [2]: df = pd.DataFrame({'A': [1, 2, np.nan],  
                           'B': [5, np.nan, np.nan],  
                           'C': [1, 2, 3]})
```

```
In [3]: df
```

Out[3]:

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

Dropna() method

```
In [4]: df.dropna()
```

Out[4]:

	A	B	C
0	1.0	5.0	1

```
In [5]: df.dropna(axis=1)
```

Out[5]:

	C
0	1
1	2
2	3

```
In [6]: df.dropna(thresh=2)
```

Out[6]:

	A	B	C
0	1.0	5.0	1
1	2.0	NaN	2

Fillna() method

```
In [7]: df.fillna(value='FILL VALUE')
```

```
Out[7]:
```

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

```
In [10]: df['A'].fillna(value=df['A'].mean())
```

```
Out[10]: 0    1.0  
         1    2.0  
         2    1.5  
         Name: A, dtype: float64
```

```
In [11]: df
```

```
Out[11]:
```

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