

SIT719 Task 2.1 Sanket Thakur

December 5, 2020

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

```
[4]: iris_data = pd.read_csv("https://archive.ics.uci.edu/ml/
↳machine-learning-databases/iris/iris.data", header=None)
```

```
[5]: iris_data
```

```
[5]:      0      1      2      3      4
0    5.1  3.5  1.4  0.2    Iris-setosa
1    4.9  3.0  1.4  0.2    Iris-setosa
2    4.7  3.2  1.3  0.2    Iris-setosa
3    4.6  3.1  1.5  0.2    Iris-setosa
4    5.0  3.6  1.4  0.2    Iris-setosa
..    ...    ...    ...    ...    ...
145   6.7  3.0  5.2  2.3  Iris-virginica
146   6.3  2.5  5.0  1.9  Iris-virginica
147   6.5  3.0  5.2  2.0  Iris-virginica
148   6.2  3.4  5.4  2.3  Iris-virginica
149   5.9  3.0  5.1  1.8  Iris-virginica
```

[150 rows x 5 columns]

```
[8]: iris_data.
↳columns=['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'class']
```

```
[9]: iris_data
```

```
[9]:      sepal_length  sepal_width  petal_length  petal_width      class
0              5.1           3.5           1.4           0.2  Iris-setosa
1              4.9           3.0           1.4           0.2  Iris-setosa
2              4.7           3.2           1.3           0.2  Iris-setosa
3              4.6           3.1           1.5           0.2  Iris-setosa
4              5.0           3.6           1.4           0.2  Iris-setosa
..              ...           ...           ...           ...    ...
145             6.7           3.0           5.2           2.3  Iris-virginica
146             6.3           2.5           5.0           1.9  Iris-virginica
```

147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

[150 rows x 5 columns]

```
[10]: iris_data.isnull().tail()
```

```
[10]:      sepal_length  sepal_width  petal_length  petal_width  class
145          False          False          False          False  False
146          False          False          False          False  False
147          False          False          False          False  False
148          False          False          False          False  False
149          False          False          False          False  False
```

```
[11]: iris_data.loc[9:29, 'petal_length'] = pd.np.NaN
```

<ipython-input-11-9f54047bc343>:1: FutureWarning: The pandas.np module is deprecated and will be removed from pandas in a future version. Import numpy directly instead

```
iris_data.loc[9:29, 'petal_length'] = pd.np.NaN
```

```
[13]: iris_data.isnull()
```

```
[13]:      sepal_length  sepal_width  petal_length  petal_width  class
0          False          False          False          False  False
1          False          False          False          False  False
2          False          False          False          False  False
3          False          False          False          False  False
4          False          False          False          False  False
..          ...          ...          ...          ...          ...
145         False          False          False          False  False
146         False          False          False          False  False
147         False          False          False          False  False
148         False          False          False          False  False
149         False          False          False          False  False
```

[150 rows x 5 columns]

```
[14]: pd.isnull(iris_data).sum().sum()
```

```
[14]: 21
```

```
[15]: iris_data.fillna(10.0)
```

```
[15]:      sepal_length  sepal_width  petal_length  petal_width  class
0          5.1          3.5          1.4          0.2  Iris-setosa
```

1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
..
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

[150 rows x 5 columns]

[]: