pandas3

September 16, 2020

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

```
In [2]: import pandas as pd
```

1 read_csv --> read the content which avaliable in csv format

```
In [4]: iris_data = pd.read_csv('C:\\Users\\admin\\Desktop\\pandas2\\iris.csv')
```

```
In [7]: #print(iris_data)
```

```
In [8]: iris_data10 = pd.read_csv('C:/Users/admin/Desktop/pandas2/iris.csv')
```

```
In [10]: #print(iris_data10)
```

```
In [11]: iris_data.head()
```

```
Out[11]:    sepal_length  sepal_width  petal_length  petal_width species
         0           5.1          3.5           1.4          0.2  setosa
         1           4.9          3.0           1.4          0.2  setosa
         2           4.7          3.2           1.3          0.2  setosa
         3           4.6          3.1           1.5          0.2  setosa
         4           5.0          3.6           1.4          0.2  setosa
```

```
In [12]: iris_data.tail()
```

```
Out[12]:      sepal_length  sepal_width  petal_length  petal_width    species
         145           6.7          3.0           5.2          2.3  virginica
         146           6.3          2.5           5.0          1.9  virginica
         147           6.5          3.0           5.2          2.0  virginica
         148           6.2          3.4           5.4          2.3  virginica
         149           5.9          3.0           5.1          1.8  virginica
```

```
In [13]: iris_data.tail(2)
```

```
Out[13]:      sepal_length  sepal_width  petal_length  petal_width    species
         148           6.2          3.4           5.4          2.3  virginica
         149           5.9          3.0           5.1          1.8  virginica
```

```
In [14]: iris_data.head(2)
```

```
Out[14]:    sepal_length  sepal_width  petal_length  petal_width species
         0           5.1          3.5           1.4          0.2  setosa
         1           4.9          3.0           1.4          0.2  setosa

In [15]: iris_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
sepal_length    150 non-null float64
sepal_width     150 non-null float64
petal_length    150 non-null float64
petal_width     150 non-null float64
species         150 non-null object
dtypes: float64(4), object(1)
memory usage: 5.9+ KB


In [16]: iris_data.shape

Out[16]: (150, 5)

In [17]: # i want remove the duplicate values

In [18]: iris_data = iris_data.append(iris_data)

In [19]: iris_data.shape

Out[19]: (300, 5)

In [20]: # dropping the duplicate values

In [21]: iris_data = iris_data.drop_duplicates()

In [22]: iris_data.shape

Out[22]: (147, 5)

In [23]: iris_data.to_csv('C:\\Users\\admin\\Desktop\\pandas2\\iris10.csv')
```