pandas_10min

January 11, 2023

1 *Day-10*

1.1 Subject: Pandas in 10 minutes on Iris Dataset

Author: Kaleem Ullah
Date: 2023-01-10
Email: kaleemrao417@outlook.com

1.2 Contents

1.2.1 Q. What is Iris

Ans.
iris, (genus Iris), genus of about 300 species of plants in the family Iridaceae, including some of the world's most popular and varied garden flowers. The diversity of the genus is centred in the north temperate zone, though some of its most handsome species are native to the Mediterranean and central Asian areas.

1.2.2 Q. What is Iris Dataset

Ans. * The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician and biologist **Ronald Fisher** in his **1936** paper The use of multiple measurements in taxonomic problems as an example of linear discriminant analysis. * The data set consists of 50 samples from each of three species of Iris **(Iris setosa, Iris virginica and Iris versicolor)**. Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

1.2.3 Q. My aim in this notebook

Ans.
I am going to use Pandas to analyze the Iris Dataset following 10 minutes tutorial of Pandas

1.2.4 Step-1: Import Libraries and Dataset

```
[ ]: import numpy as np
     import pandas as pd
     import seaborn as sns
     import warnings
     warnings.filterwarnings('ignore')
     df = sns.load_dataset('iris')
```

1.2.5 Step-2: See the Basic section

```
[ ]: df.head()
```

```
[ ]:    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
```

```
[ ]: df.tail()
```

```
[ ]:      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
```

1.2.6 Step-3: Columns in data

```
[ ]: df.columns
```

```
[ ]: Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
            'species'],
           dtype='object')
```

1.2.7 Step-4: Convert Dataframe into Numpy Arrays

```
[ ]: df.to_numpy()
```

```
[ ]: array([[5.1, 3.5, 1.4, 0.2, 'setosa'],
            [4.9, 3.0, 1.4, 0.2, 'setosa'],
            [4.7, 3.2, 1.3, 0.2, 'setosa'],
            [4.6, 3.1, 1.5, 0.2, 'setosa'],
            [5.0, 3.6, 1.4, 0.2, 'setosa'],
            [5.4, 3.9, 1.7, 0.4, 'setosa'],
            [4.6, 3.4, 1.4, 0.3, 'setosa'],
            [5.0, 3.4, 1.5, 0.2, 'setosa'],
            [4.4, 2.9, 1.4, 0.2, 'setosa'],
            [4.9, 3.1, 1.5, 0.1, 'setosa'],
            [5.4, 3.7, 1.5, 0.2, 'setosa'],
            [4.8, 3.4, 1.6, 0.2, 'setosa'],
            [4.8, 3.0, 1.4, 0.1, 'setosa'],
            [4.3, 3.0, 1.1, 0.1, 'setosa'],
            [5.8, 4.0, 1.2, 0.2, 'setosa'],
            [5.7, 4.4, 1.5, 0.4, 'setosa'],
            [5.4, 3.9, 1.3, 0.4, 'setosa'],
            [5.1, 3.5, 1.4, 0.3, 'setosa'],
            [5.7, 3.8, 1.7, 0.3, 'setosa'],
            [5.1, 3.8, 1.5, 0.3, 'setosa'],
            [5.4, 3.4, 1.7, 0.2, 'setosa'],
            [5.1, 3.7, 1.5, 0.4, 'setosa'],
            [4.6, 3.6, 1.0, 0.2, 'setosa'],
            [5.1, 3.3, 1.7, 0.5, 'setosa'],
            [4.8, 3.4, 1.9, 0.2, 'setosa'],
            [5.0, 3.0, 1.6, 0.2, 'setosa'],
            [5.0, 3.4, 1.6, 0.4, 'setosa'],
            [5.2, 3.5, 1.5, 0.2, 'setosa'],
            [5.2, 3.4, 1.4, 0.2, 'setosa'],
            [4.7, 3.2, 1.6, 0.2, 'setosa'],
            [4.8, 3.1, 1.6, 0.2, 'setosa'],
            [5.4, 3.4, 1.5, 0.4, 'setosa'],
            [5.2, 4.1, 1.5, 0.1, 'setosa'],
            [5.5, 4.2, 1.4, 0.2, 'setosa'],
            [4.9, 3.1, 1.5, 0.2, 'setosa'],
            [5.0, 3.2, 1.2, 0.2, 'setosa'],
            [5.5, 3.5, 1.3, 0.2, 'setosa'],
            [4.9, 3.6, 1.4, 0.1, 'setosa'],
            [4.4, 3.0, 1.3, 0.2, 'setosa'],
            [5.1, 3.4, 1.5, 0.2, 'setosa'],
            [5.0, 3.5, 1.3, 0.3, 'setosa'],
            [4.5, 2.3, 1.3, 0.3, 'setosa'],
            [4.4, 3.2, 1.3, 0.2, 'setosa'],
            [5.0, 3.5, 1.6, 0.6, 'setosa'],
            [5.1, 3.8, 1.9, 0.4, 'setosa'],
            [4.8, 3.0, 1.4, 0.3, 'setosa'],
            [5.1, 3.8, 1.6, 0.2, 'setosa'],
            [4.6, 3.2, 1.4, 0.2, 'setosa'],
            [5.3, 3.7, 1.5, 0.2, 'setosa'],
            [5.0, 3.3, 1.4, 0.2, 'setosa'],
            [7.0, 3.2, 4.7, 1.4, 'versicolor'],
            [6.4, 3.2, 4.5, 1.5, 'versicolor'],
            [6.9, 3.1, 4.9, 1.5, 'versicolor'],
            [5.5, 2.3, 4.0, 1.3, 'versicolor'],
            [6.5, 2.8, 4.6, 1.5, 'versicolor'],
            [5.7, 2.8, 4.5, 1.3, 'versicolor'],
            [6.3, 3.3, 4.7, 1.6, 'versicolor'],
            [4.9, 2.4, 3.3, 1.0, 'versicolor'],
            [6.6, 2.9, 4.6, 1.3, 'versicolor'],
            [5.2, 2.7, 3.9, 1.4, 'versicolor'],
            [5.0, 2.0, 3.5, 1.0, 'versicolor'],
            [5.9, 3.0, 4.2, 1.5, 'versicolor'],
            [6.0, 2.2, 4.0, 1.0, 'versicolor'],
            [6.1, 2.9, 4.7, 1.4, 'versicolor'],
            [5.6, 2.9, 3.6, 1.3, 'versicolor'],
            [6.7, 3.1, 4.4, 1.4, 'versicolor'],
            [5.6, 3.0, 4.5, 1.5, 'versicolor'],
            [5.8, 2.7, 4.1, 1.0, 'versicolor'],
            [6.2, 2.2, 4.5, 1.5, 'versicolor'],
            [5.6, 2.5, 3.9, 1.1, 'versicolor'],
            [5.9, 3.2, 4.8, 1.8, 'versicolor'],
            [6.1, 2.8, 4.0, 1.3, 'versicolor'],
            [6.3, 2.5, 4.9, 1.5, 'versicolor'],
            [6.1, 2.8, 4.7, 1.2, 'versicolor'],
            [6.4, 2.9, 4.3, 1.3, 'versicolor'],
            [6.6, 3.0, 4.4, 1.4, 'versicolor'],
            [6.8, 2.8, 4.8, 1.4, 'versicolor'],
            [6.7, 3.0, 5.0, 1.7, 'versicolor'],
            [6.0, 2.9, 4.5, 1.5, 'versicolor'],
            [5.7, 2.6, 3.5, 1.0, 'versicolor'],
            [5.5, 2.4, 3.8, 1.1, 'versicolor'],
            [5.5, 2.4, 3.7, 1.0, 'versicolor'],
            [5.8, 2.7, 3.9, 1.2, 'versicolor'],
            [6.0, 2.7, 5.1, 1.6, 'versicolor'],
            [5.4, 3.0, 4.5, 1.5, 'versicolor'],
            [6.0, 3.4, 4.5, 1.6, 'versicolor'],
            [6.7, 3.1, 4.7, 1.5, 'versicolor'],
            [6.3, 2.3, 4.4, 1.3, 'versicolor'],
            [5.6, 3.0, 4.1, 1.3, 'versicolor'],
            [5.5, 2.5, 4.0, 1.3, 'versicolor'],
            [5.5, 2.6, 4.4, 1.2, 'versicolor'],
            [6.1, 3.0, 4.6, 1.4, 'versicolor'],
            [5.8, 2.6, 4.0, 1.2, 'versicolor'],
            [5.0, 2.3, 3.3, 1.0, 'versicolor'],
            [5.6, 2.7, 4.2, 1.3, 'versicolor'],
            [5.7, 3.0, 4.2, 1.2, 'versicolor'],
            [5.7, 2.9, 4.2, 1.3, 'versicolor'],
            [6.2, 2.9, 4.3, 1.3, 'versicolor'],
            [5.1, 2.5, 3.0, 1.1, 'versicolor'],
            [5.7, 2.8, 4.1, 1.3, 'versicolor'],
            [6.3, 3.3, 6.0, 2.5, 'virginica'],
            [5.8, 2.7, 5.1, 1.9, 'virginica'],
            [7.1, 3.0, 5.9, 2.1, 'virginica'],
            [6.3, 2.9, 5.6, 1.8, 'virginica'],
            [6.5, 3.0, 5.8, 2.2, 'virginica'],
            [7.6, 3.0, 6.6, 2.1, 'virginica'],
            [4.9, 2.5, 4.5, 1.7, 'virginica'],
            [7.3, 2.9, 6.3, 1.8, 'virginica'],
            [6.7, 2.5, 5.8, 1.8, 'virginica'],
            [7.2, 3.6, 6.1, 2.5, 'virginica'],
            [6.5, 3.2, 5.1, 2.0, 'virginica'],
            [6.4, 2.7, 5.3, 1.9, 'virginica'],
            [6.8, 3.0, 5.5, 2.1, 'virginica'],
            [5.7, 2.5, 5.0, 2.0, 'virginica'],
            [5.8, 2.8, 5.1, 2.4, 'virginica'],
            [6.4, 3.2, 5.3, 2.3, 'virginica'],
            [6.5, 3.0, 5.5, 1.8, 'virginica'],
            [7.7, 3.8, 6.7, 2.2, 'virginica'],
            [7.7, 2.6, 6.9, 2.3, 'virginica'],
            [6.0, 2.2, 5.0, 1.5, 'virginica'],
            [6.9, 3.2, 5.7, 2.3, 'virginica'],
            [5.6, 2.8, 4.9, 2.0, 'virginica'],
            [7.7, 2.8, 6.7, 2.0, 'virginica'],
            [6.3, 2.7, 4.9, 1.8, 'virginica'],
            [6.7, 3.3, 5.7, 2.1, 'virginica'],
            [7.2, 3.2, 6.0, 1.8, 'virginica'],
            [6.2, 2.8, 4.8, 1.8, 'virginica'],
            [6.1, 3.0, 4.9, 1.8, 'virginica'],
            [6.4, 2.8, 5.6, 2.1, 'virginica'],
            [7.2, 3.0, 5.8, 1.6, 'virginica'],
            [7.4, 2.8, 6.1, 1.9, 'virginica'],
            [7.9, 3.8, 6.4, 2.0, 'virginica'],
            [6.4, 2.8, 5.6, 2.2, 'virginica'],
            [6.3, 2.8, 5.1, 1.5, 'virginica'],
            [6.1, 2.6, 5.6, 1.4, 'virginica'],
            [7.7, 3.0, 6.1, 2.3, 'virginica'],
            [6.3, 3.4, 5.6, 2.4, 'virginica'],
            [6.4, 3.1, 5.5, 1.8, 'virginica'],
            [6.0, 3.0, 4.8, 1.8, 'virginica'],
            [6.9, 3.1, 5.4, 2.1, 'virginica'],
            [6.7, 3.1, 5.6, 2.4, 'virginica'],
            [6.9, 3.1, 5.1, 2.3, 'virginica'],
            [5.8, 2.7, 5.1, 1.9, 'virginica'],
            [6.8, 3.2, 5.9, 2.3, 'virginica'],
            [6.7, 3.3, 5.7, 2.5, 'virginica'],
            [6.7, 3.0, 5.2, 2.3, 'virginica'],
            [6.3, 2.5, 5.0, 1.9, 'virginica'],
            [6.5, 3.0, 5.2, 2.0, 'virginica'],
            [6.2, 3.4, 5.4, 2.3, 'virginica'],
            [5.9, 3.0, 5.1, 1.8, 'virginica']], dtype=object)
```

1.2.8 Step-5: Type of dataframe

[ ]:
```
df.dtypes
```

[ ]:
```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species          object
dtype: object
```

1.2.9 Step-6: Shape of dataframe

[ ]:
```
df.shape
```

[ ]:
```
(150, 5)
```

1.2.10 Step-7: Info of dataframe

[ ]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
```

```
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

1.2.11 Step-8: Stats in dataframe

```
[ ]: df.describe()
```

```
[ ]:        sepal_length  sepal_width  petal_length  petal_width
     count    150.000000   150.000000    150.000000   150.000000
     mean       5.843333     3.057333      3.758000     1.199333
     std        0.828066     0.435866      1.765298     0.762238
     min        4.300000     2.000000      1.000000     0.100000
     25%        5.100000     2.800000      1.600000     0.300000
     50%        5.800000     3.000000      4.350000     1.300000
     75%        6.400000     3.300000      5.100000     1.800000
     max        7.900000     4.400000      6.900000     2.500000
```

1.2.12 Step-9: Transpose the dataframe

```
[ ]: df.T
```

```
[ ]:                      0       1       2       3       4       5       6       7    \
     sepal_length       5.1     4.9     4.7     4.6     5.0     5.4     4.6     5.0
     sepal_width        3.5     3.0     3.2     3.1     3.6     3.9     3.4     3.4
     petal_length       1.4     1.4     1.3     1.5     1.4     1.7     1.4     1.5
     petal_width        0.2     0.2     0.2     0.2     0.2     0.4     0.3     0.2
     species         setosa  setosa  setosa  setosa  setosa  setosa  setosa  setosa

                         8       9    …        140        141        142        143  \
     sepal_length       4.4     4.9   …        6.7        6.9        5.8        6.8
     sepal_width        2.9     3.1   …        3.1        3.1        2.7        3.2
     petal_length       1.4     1.5   …        5.6        5.1        5.1        5.9
     petal_width        0.2     0.1   …        2.4        2.3        1.9        2.3
     species         setosa  setosa   …  virginica  virginica  virginica  virginica

                           144        145        146        147        148        149
     sepal_length          6.7        6.7        6.3        6.5        6.2        5.9
     sepal_width           3.3        3.0        2.5        3.0        3.4        3.0
     petal_length          5.7        5.2        5.0        5.2        5.4        5.1
```

```
petal_width         2.5        2.3        1.9        2.0        2.3        1.8
species       virginica  virginica  virginica  virginica  virginica  virginica

[5 rows x 150 columns]
```

1.2.13 Step-10: Sort by axis

- **axis=1** means sort by columns

```
[ ]: df.sort_index(axis=1, ascending=False)
```

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

     [150 rows x 5 columns]
```

- **axis=0** means sort by row

```
[ ]: df.sort_index(axis=0, ascending=False)
```

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

     [150 rows x 5 columns]
```

1.2.14 Step-11: Sort by values

[ ]:
```
df.sort_values(by='sepal_length', ascending=False)
```

[ ]:
```
     sepal_length  sepal_width  petal_length  petal_width    species
131           7.9          3.8           6.4          2.0  virginica
135           7.7          3.0           6.1          2.3  virginica
122           7.7          2.8           6.7          2.0  virginica
117           7.7          3.8           6.7          2.2  virginica
118           7.7          2.6           6.9          2.3  virginica
..              …            …             …            …          …
41            4.5          2.3           1.3          0.3     setosa
42            4.4          3.2           1.3          0.2     setosa
38            4.4          3.0           1.3          0.2     setosa
8             4.4          2.9           1.4          0.2     setosa
13            4.3          3.0           1.1          0.1     setosa

[150 rows x 5 columns]
```

1.2.15 Step-12: Selecting a single column

[ ]:
```
df['sepal_length']
```

[ ]:
```
0      5.1
1      4.9
2      4.7
3      4.6
4      5.0
      …
145    6.7
146    6.3
147    6.5
148    6.2
149    5.9
Name: sepal_length, Length: 150, dtype: float64
```

1.2.16 Step-13: Selecting Multiple Columns

[ ]:
```
df[['sepal_length', 'sepal_width']]
```

[ ]:
```
     sepal_length  sepal_width
0             5.1          3.5
1             4.9          3.0
2             4.7          3.2
3             4.6          3.1
4             5.0          3.6
..              …            …
145           6.7          3.0
```

```
146           6.3          2.5
147           6.5          3.0
148           6.2          3.4
149           5.9          3.0

[150 rows x 2 columns]
```

1.2.17 Step-14: Slicing rows

```
[ ]: df[22:29]
```

```
[ ]:     sepal_length  sepal_width  petal_length  petal_width species
     22           4.6          3.6           1.0          0.2  setosa
     23           5.1          3.3           1.7          0.5  setosa
     24           4.8          3.4           1.9          0.2  setosa
     25           5.0          3.0           1.6          0.2  setosa
     26           5.0          3.4           1.6          0.4  setosa
     27           5.2          3.5           1.5          0.2  setosa
     28           5.2          3.4           1.4          0.2  setosa
```

1.2.18 Step-15: Slicing at Specific index(Position)

```
[ ]: df.iloc[27]
```

```
[ ]: sepal_length       5.2
     sepal_width        3.5
     petal_length       1.5
     petal_width        0.2
     species         setosa
     Name: 27, dtype: object
```

1.2.19 Step-16: Slicing multiple rows and column

```
[ ]: df.iloc[26:32, 0:3]
```

```
[ ]:     sepal_length  sepal_width  petal_length
     26           5.0          3.4           1.6
     27           5.2          3.5           1.5
     28           5.2          3.4           1.4
     29           4.7          3.2           1.6
     30           4.8          3.1           1.6
     31           5.4          3.4           1.5
```

1.2.20 Step-17: Boolian indexing

```
[ ]: df[df['sepal_length'] > 7]
[ ]:      sepal_length  sepal_width  petal_length  petal_width    species
     102           7.1          3.0           5.9          2.1  virginica
     105           7.6          3.0           6.6          2.1  virginica
     107           7.3          2.9           6.3          1.8  virginica
     109           7.2          3.6           6.1          2.5  virginica
     117           7.7          3.8           6.7          2.2  virginica
     118           7.7          2.6           6.9          2.3  virginica
     122           7.7          2.8           6.7          2.0  virginica
     125           7.2          3.2           6.0          1.8  virginica
     129           7.2          3.0           5.8          1.6  virginica
     130           7.4          2.8           6.1          1.9  virginica
     131           7.9          3.8           6.4          2.0  virginica
     135           7.7          3.0           6.1          2.3  virginica
```

```
[ ]: df.iloc[:, 0:4] > 3
```

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

     [150 rows x 4 columns]
```

1.2.21 Step-18: Filtering by isin( )

```
[ ]: df[df['species'].isin(['setosa', 'virginica'])]
```

```
[ ]:      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
     ..            …           …            …            …          …
     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
```

```
[100 rows x 5 columns]
```

1.2.22 Step-19: Setting

- **Setting a new column automatically aligns the data by the indexes**

[ ]:

```
df['new_column'] = np.arange(150)
df.head()
```

[ ]:

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

- Setting values by label

[ ]:

```
df.iat[3, 5] = 100                                # iat specify only 1 position
df.head()                                         # unlike iloc, which specify a range
```

[ ]:

```
   sepal_length  sepal_width  petal_length  petal_width species  new_column
0           5.1          3.5           1.4          0.2  setosa           0
1           4.9          3.0           1.4          0.2  setosa           1
2           4.7          3.2           1.3          0.2  setosa           2
3           4.6          3.1           1.5          0.2  setosa         100
4           5.0          3.6           1.4          0.2  setosa           4
```

1.2.23 Step-20: Creating Missing Data

[ ]:

```
df.loc[0:3, 'sepal_length'] = np.nan; df.loc[4:7, 'sepal_width'] = np.nan
df.head(8)
```

[ ]:

```
   sepal_length  sepal_width  petal_length  petal_width species  new_column
0           NaN          3.5           1.4          0.2  setosa           0
1           NaN          3.0           1.4          0.2  setosa           1
2           NaN          3.2           1.3          0.2  setosa           2
3           NaN          3.1           1.5          0.2  setosa         100
4           5.0          NaN           1.4          0.2  setosa           4
5           5.4          NaN           1.7          0.4  setosa           5
6           4.6          NaN           1.4          0.3  setosa           6
7           5.0          NaN           1.5          0.2  setosa           7
```

- **Checking Missing Values**

[ ]:

```
df.isnull().sum()
```

```
[ ]: sepal_length    4
     sepal_width     4
     petal_length    0
     petal_width     0
     species         0
     new_column      0
     dtype: int64
```

1.2.24 Step-21: Dropping Missing Data

```
[ ]: df.dropna(axis=0, inplace=True)
     df.head()
```

```
[ ]:     sepal_length  sepal_width  petal_length  petal_width species  new_column
     8            4.4          2.9           1.4          0.2  setosa           8
     9            4.9          3.1           1.5          0.1  setosa           9
     10           5.4          3.7           1.5          0.2  setosa          10
     11           4.8          3.4           1.6          0.2  setosa          11
     12           4.8          3.0           1.4          0.1  setosa          12
```

1.2.25 Step-22: Reset-Index after dropping NaN

```
[ ]: df.reset_index(drop=True, inplace=True)
     df.head()
```

```
[ ]:    sepal_length  sepal_width  petal_length  petal_width species  new_column
     0           4.4          2.9           1.4          0.2  setosa           8
     1           4.9          3.1           1.5          0.1  setosa           9
     2           5.4          3.7           1.5          0.2  setosa          10
     3           4.8          3.4           1.6          0.2  setosa          11
     4           4.8          3.0           1.4          0.1  setosa          12
```

```
[ ]: df.tail()
```

```
[ ]:      sepal_length  sepal_width  petal_length  petal_width    species  \
     137           6.7          3.0           5.2          2.3  virginica
     138           6.3          2.5           5.0          1.9  virginica
     139           6.5          3.0           5.2          2.0  virginica
     140           6.2          3.4           5.4          2.3  virginica
     141           5.9          3.0           5.1          1.8  virginica

          new_column
     137         145
     138         146
     139         147
     140         148
     141         149
```

1.2.26 Step-23: Filling Missing Data

- **Create Missing Values**

```
[ ]: df.loc[0:3, 'sepal_length'] = np.nan; df.loc[4:7, 'sepal_width'] = np.nan
     df.head(8)
```

```
[ ]:    sepal_length  sepal_width  petal_length  petal_width species  new_column
     0           NaN          2.9           1.4          0.2  setosa           8
     1           NaN          3.1           1.5          0.1  setosa           9
     2           NaN          3.7           1.5          0.2  setosa          10
     3           NaN          3.4           1.6          0.2  setosa          11
     4           4.8          NaN           1.4          0.1  setosa          12
     5           4.3          NaN           1.1          0.1  setosa          13
     6           5.8          NaN           1.2          0.2  setosa          14
     7           5.7          NaN           1.5          0.4  setosa          15
```

- **Filling Missing Data with Mean**

```
[ ]: df.describe()
```

```
[ ]:        sepal_length  sepal_width  petal_length  petal_width  new_column
     count    138.000000   138.000000    142.000000   142.000000  142.000000
     mean       5.925362     3.022464      3.888028     1.253521   78.500000
     std        0.808356     0.418377      1.724273     0.747161   41.135953
     min        4.300000     2.000000      1.000000     0.100000    8.000000
     25%        5.225000     2.800000      1.600000     0.400000   43.250000
     50%        5.900000     3.000000      4.450000     1.400000   78.500000
     75%        6.475000     3.300000      5.100000     1.800000  113.750000
     max        7.900000     4.200000      6.900000     2.500000  149.000000
```

```
[ ]: df.fillna(df.mean(), inplace=True)
     df.head(8)
```

```
[ ]:    sepal_length  sepal_width  petal_length  petal_width species  new_column
     0      5.925362     2.900000           1.4          0.2  setosa           8
     1      5.925362     3.100000           1.5          0.1  setosa           9
     2      5.925362     3.700000           1.5          0.2  setosa          10
     3      5.925362     3.400000           1.6          0.2  setosa          11
     4      4.800000     3.022464           1.4          0.1  setosa          12
     5      4.300000     3.022464           1.1          0.1  setosa          13
     6      5.800000     3.022464           1.2          0.2  setosa          14
     7      5.700000     3.022464           1.5          0.4  setosa          15
```

```
[ ]:
```