Iris

Introduction:

This exercise may seem a little bit strange, but keep doing it.

Step 1. Import the necessary libraries

In [30]:
```
import pandas as pd
import numpy as np
```

Step 2. Import the dataset from this address.

In [31]:
```
pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data",sep=',')
```

Out[31]:

	5.1	3.5	1.4	0.2	Iris-setosa
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa
...	...	...	...	...	...
144	6.7	3.0	5.2	2.3	Iris-virginica
145	6.3	2.5	5.0	1.9	Iris-virginica
146	6.5	3.0	5.2	2.0	Iris-virginica
147	6.2	3.4	5.4	2.3	Iris-virginica
148	5.9	3.0	5.1	1.8	Iris-virginica

149 rows × 5 columns

Step 3. Assign it to a variable called iris

In [32]:
```
iris=pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data",sep=',')
iris.head()
```

Out[32]:

	5.1	3.5	1.4	0.2	Iris-setosa
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa

Step 4. Create columns for the dataset

In [84]:
```
# 1. sepal_length (in cm)
# 2. sepal_width (in cm)
# 3. petal_length (in cm)
# 4. petal_width (in cm)
# 5. class
```

In [33]:
```
list1=['sepal_length (in cm)','sepal_width (in cm)','petal_length (in cm)','petal_width (in cm)','class']
list1
```

Out[33]:
```
['sepal_length (in cm)',
 'sepal_width (in cm)',
 'petal_length (in cm)',
 'petal_width (in cm)',
 'class']
```

In [34]:
```
iris.columns=list1
iris.head()
```

Out[34]:

	sepal_length (in cm)	sepal_width (in cm)	petal_length (in cm)	petal_width (in cm)	class
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa

Step 5. Is there any missing value in the dataframe?

In [35]:
```
iris.isnull().sum()
```

Out[35]:
```
sepal_length (in cm)    0
sepal_width (in cm)     0
petal_length (in cm)    0
petal_width (in cm)     0
class                   0
dtype: int64
```

Step 6. Lets set the values of the rows 10 to 29 of the column 'petal_length' to NaN

In [25]:
```
#1st Method
iris['petal_length (in cm)'][10:30] = np.nan
```

```
/tmp/ipykernel_32/2172142831.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy
  iris['petal_length (in cm)'][10:30] = np.nan
```

In [26]:
```
iris.iloc[10:30]
```

Out[26]:

	sepal_length (in cm)	sepal_width (in cm)	petal_length (in cm)	petal_width (in cm)	class
10	4.8	3.4	NaN	0.2	Iris-setosa
11	4.8	3.0	NaN	0.1	Iris-setosa
12	4.3	3.0	NaN	0.1	Iris-setosa
13	5.8	4.0	NaN	0.2	Iris-setosa
14	5.7	4.4	NaN	0.4	Iris-setosa
15	5.4	3.9	NaN	0.4	Iris-setosa
16	5.1	3.5	NaN	0.3	Iris-setosa
17	5.7	3.8	NaN	0.3	Iris-setosa
18	5.1	3.8	NaN	0.3	Iris-setosa
19	5.4	3.4	NaN	0.2	Iris-setosa
20	5.1	3.7	NaN	0.4	Iris-setosa
21	4.6	3.6	NaN	0.2	Iris-setosa
22	5.1	3.3	NaN	0.5	Iris-setosa
23	4.8	3.4	NaN	0.2	Iris-setosa
24	5.0	3.0	NaN	0.2	Iris-setosa
25	5.0	3.4	NaN	0.4	Iris-setosa
26	5.2	3.5	NaN	0.2	Iris-setosa
27	5.2	3.4	NaN	0.2	Iris-setosa
28	4.7	3.2	NaN	0.2	Iris-setosa
29	4.8	3.1	NaN	0.2	Iris-setosa

In [36]:
```
#2nd Method
iris.iloc[10:30,2]=np.nan
iris.iloc[10:30,0:5]
```

Out[36]:

	sepal_length (in cm)	sepal_width (in cm)	petal_length (in cm)	petal_width (in cm)	class
10	4.8	3.4	NaN	0.2	Iris-setosa
11	4.8	3.0	NaN	0.1	Iris-setosa
12	4.3	3.0	NaN	0.1	Iris-setosa
13	5.8	4.0	NaN	0.2	Iris-setosa
14	5.7	4.4	NaN	0.4	Iris-setosa
15	5.4	3.9	NaN	0.4	Iris-setosa
16	5.1	3.5	NaN	0.3	Iris-setosa
17	5.7	3.8	NaN	0.3	Iris-setosa
18	5.1	3.8	NaN	0.3	Iris-setosa
19	5.4	3.4	NaN	0.2	Iris-setosa
20	5.1	3.7	NaN	0.4	Iris-setosa
21	4.6	3.6	NaN	0.2	Iris-setosa
22	5.1	3.3	NaN	0.5	Iris-setosa
23	4.8	3.4	NaN	0.2	Iris-setosa
24	5.0	3.0	NaN	0.2	Iris-setosa
25	5.0	3.4	NaN	0.4	Iris-setosa
26	5.2	3.5	NaN	0.2	Iris-setosa
27	5.2	3.4	NaN	0.2	Iris-setosa
28	4.7	3.2	NaN	0.2	Iris-setosa
29	4.8	3.1	NaN	0.2	Iris-setosa

In [37]:

```
iris.isna().sum()
```

Out[37]:

```
sepal_length (in cm)     0
sepal_width (in cm)      0
petal_length (in cm)    20
petal_width (in cm)      0
class                    0
dtype: int64
```

Step 7. Good, now lets substitute the NaN values to 1.0

In [38]:

```
iris.iloc[10:30,2].fillna(1.0,inplace = True)
iris.iloc[10:30,0:5]
```

Out[38]:

	sepal_length (in cm)	sepal_width (in cm)	petal_length (in cm)	petal_width (in cm)	class
10	4.8	3.4	1.0	0.2	Iris-setosa
11	4.8	3.0	1.0	0.1	Iris-setosa
12	4.3	3.0	1.0	0.1	Iris-setosa
13	5.8	4.0	1.0	0.2	Iris-setosa
14	5.7	4.4	1.0	0.4	Iris-setosa
15	5.4	3.9	1.0	0.4	Iris-setosa
16	5.1	3.5	1.0	0.3	Iris-setosa
17	5.7	3.8	1.0	0.3	Iris-setosa
18	5.1	3.8	1.0	0.3	Iris-setosa
19	5.4	3.4	1.0	0.2	Iris-setosa
20	5.1	3.7	1.0	0.4	Iris-setosa
21	4.6	3.6	1.0	0.2	Iris-setosa
22	5.1	3.3	1.0	0.5	Iris-setosa
23	4.8	3.4	1.0	0.2	Iris-setosa
24	5.0	3.0	1.0	0.2	Iris-setosa
25	5.0	3.4	1.0	0.4	Iris-setosa
26	5.2	3.5	1.0	0.2	Iris-setosa
27	5.2	3.4	1.0	0.2	Iris-setosa
28	4.7	3.2	1.0	0.2	Iris-setosa
29	4.8	3.1	1.0	0.2	Iris-setosa

Step 8. Now let's delete the column class

In [39]:

```
iris.drop(labels='class', axis=1)
```

Out[39]:

	sepal_length (in cm)	sepal_width (in cm)	petal_length (in cm)	petal_width (in cm)
0	4.9	3.0	1.4	0.2
1	4.7	3.2	1.3	0.2
2	4.6	3.1	1.5	0.2
3	5.0	3.6	1.4	0.2
4	5.4	3.9	1.7	0.4
...	...	...	...	...
144	6.7	3.0	5.2	2.3
145	6.3	2.5	5.0	1.9
146	6.5	3.0	5.2	2.0
147	6.2	3.4	5.4	2.3
148	5.9	3.0	5.1	1.8

149 rows × 4 columns

Step 9. Set the first 3 rows as NaN

In [42]:

```
iris.iloc[0:3,:]=np.nan
iris.iloc[0:3,:]
```

Out[42]:

	sepal_length (in cm)	sepal_width (in cm)	petal_length (in cm)	petal_width (in cm)	class
0	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN

In [43]:

```
iris.isna().sum()
```

Out[43]:

```
sepal_length (in cm)    3
sepal_width (in cm)     3
petal_length (in cm)    3
petal_width (in cm)     3
class                   3
dtype: int64
```

Step 10. Delete the rows that have NaN

In [44]:

```
iris=iris.dropna(axis = 0, how = "any")
iris.head(10)
```

Out[44]:

	sepal_length (in cm)	sepal_width (in cm)	petal_length (in cm)	petal_width (in cm)	class
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa
5	4.6	3.4	1.4	0.3	Iris-setosa
6	5.0	3.4	1.5	0.2	Iris-setosa
7	4.4	2.9	1.4	0.2	Iris-setosa
8	4.9	3.1	1.5	0.1	Iris-setosa
9	5.4	3.7	1.5	0.2	Iris-setosa
10	4.8	3.4	1.0	0.2	Iris-setosa
11	4.8	3.0	1.0	0.1	Iris-setosa
12	4.3	3.0	1.0	0.1	Iris-setosa

Step 11. Reset the index so it begins with 0 again

In [45]:

```
iris = iris.reset_index(drop = True)
iris.head(100)
```

Out[45]:

	sepal_length (in cm)	sepal_width (in cm)	petal_length (in cm)	petal_width (in cm)	class
0	5.0	3.6	1.4	0.2	Iris-setosa
1	5.4	3.9	1.7	0.4	Iris-setosa
2	4.6	3.4	1.4	0.3	Iris-setosa
3	5.0	3.4	1.5	0.2	Iris-setosa
4	4.4	2.9	1.4	0.2	Iris-setosa
...	...	...	...	...	...
95	5.7	2.8	4.1	1.3	Iris-versicolor
96	6.3	3.3	6.0	2.5	Iris-virginica
97	5.8	2.7	5.1	1.9	Iris-virginica
98	7.1	3.0	5.9	2.1	Iris-virginica
99	6.3	2.9	5.6	1.8	Iris-virginica

100 rows × 5 columns

BONUS: Create your own question and answer it.

In [ ]: