ASSIGNMENT SOLUTIONS:

jupyter TECHBEES.. Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3

Run Code

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
data = sns.load_dataset('iris')
print(data)
```

```
     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

[150 rows x 5 columns]
```

Print the first five records of the dataset for each 'species'

In [33]:
```
pd.concat([data.loc[(data['species']=='setosa')].head(5),data.loc[(data['species']=='versicolor')].head(5),
           data.loc[(data['species']=='virginica')].head(5)])
```

Out[33]:

	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
50	7.0	3.2	4.7	1.4	versicolor
51	6.4	3.2	4.5	1.5	versicolor
52	6.9	3.1	4.9	1.5	versicolor
53	5.5	2.3	4.0	1.3	versicolor
54	6.5	2.8	4.6	1.5	versicolor
100	6.3	3.3	6.0	2.5	virginica
101	5.8	2.7	5.1	1.9	virginica
102	7.1	3.0	5.9	2.1	virginica
103	6.3	2.9	5.6	1.8	virginica
104	6.5	3.0	5.8	2.2	virginica

Find out which sepal_length value appears the most.

In [3]:
```
data['sepal_length'].mode()
# 5.0 appears the most
```

Out[3]:
```
0    5.0
dtype: float64
```

Find out the mean sepal_width group by species.

In [28]:
```
# mean of sepal_width of setosa
data.loc[(data['species']=='setosa')].mean()['sepal_width']
```

Out[28]:
```
3.428000000000001
```

In [29]:
```
# mean of sepal_width of versicolor
data.loc[(data['species']=='versicolor')].mean()['sepal_width']
```

Out[29]:
```
2.7700000000000005
```

In [30]:
```
# mean of sepal_width of virginica
data.loc[(data['species']=='virginica')].mean()['sepal_width']
```

Out[30]:
```
2.9739999999999998
```

Find out which species has the maximum and which species has the minimum petal_length.

In [9]:
```
# virginica has maximum petal_length
data.loc[data['petal_length'].idxmax()]
```

Out[9]:
```
sepal_length          7.7
sepal_width           2.6
petal_length          6.9
petal_width           2.3
species         virginica
Name: 118, dtype: object
```

In [10]:
```
# setosa has minimum petal_length
data.loc[data['petal_length'].idxmin()]
```

Out[10]:
```
sepal_length       4.6
sepal_width        3.6
petal_length       1.0
petal_width        0.2
species         setosa
Name: 22, dtype: object
```

Find out the median petal_width for the entire dataset.

In [11]:
```
data["petal_width"].median()
# 1.3 is the median of petal_width for the entire dataset
```

Out[11]: 1.3

Print the first 20 records of the dataset, in the ascending order of sepal_length.

In [31]:
```
newdata = data.head(20)
```

In [32]:
```
newdata.sort_values(by = 'sepal_length',ascending = True)[['sepal_length']]
```

Out[32]:

	sepal_length
13	4.3
8	4.4
3	4.6
6	4.6
2	4.7
12	4.8
11	4.8
9	4.9
1	4.9
4	5.0
7	5.0
17	5.1
0	5.1
19	5.1
16	5.4
5	5.4
10	5.4
18	5.7
15	5.7
14	5.8

Print the last 20 records of the dataset, in the descending order of sepal_width.

In [14]:
```
newdata.sort_values(by = 'sepal_width',ascending = False)[['sepal_width']]
```

Out[14]:

	sepal_width
15	4.4
14	4.0
16	3.9
5	3.9
19	3.8
18	3.8
10	3.7
4	3.6
17	3.5
0	3.5
7	3.4
11	3.4
6	3.4
2	3.2
9	3.1
3	3.1
1	3.0
12	3.0
13	3.0
8	2.9

Display the species and their characteristics where the petal_length is less than 3.

```
In [19]: df = data[['petal_length','species']]
```

```
In [20]: df.loc[(df['petal_length']<3)]
```

Out[20]:

	petal_length	species
0	1.4	setosa
1	1.4	setosa
2	1.3	setosa
3	1.5	setosa
4	1.4	setosa
5	1.7	setosa
6	1.4	setosa
7	1.5	setosa
8	1.4	setosa
9	1.5	setosa
10	1.5	setosa
11	1.6	setosa
12	1.4	setosa
13	1.1	setosa
14	1.2	setosa
15	1.5	setosa
16	1.3	setosa
17	1.4	setosa
18	1.7	setosa
19	1.5	setosa
20	1.7	setosa
21	1.5	setosa
22	1.0	setosa
23	1.7	setosa
24	1.9	setosa
25	1.6	setosa
26	1.6	setosa
27	1.5	setosa
28	1.4	setosa
29	1.6	setosa
30	1.6	setosa
31	1.5	setosa
32	1.5	setosa
33	1.4	setosa
34	1.5	setosa
35	1.2	setosa
36	1.3	setosa
37	1.4	setosa
38	1.3	setosa
39	1.5	setosa
40	1.3	setosa
41	1.3	setosa
42	1.3	setosa
43	1.6	setosa
44	1.9	setosa
45	1.4	setosa
46	1.6	setosa
47	1.4	setosa
48	1.5	setosa
49	1.4	setosa

Count the number of records for the Virginica species where petal_length is more than 5.

In [22]:
```
df = data[['petal_length','species']]
```

In [23]:
```
df1 = df.loc[(df['petal_length']>5)]
```

In [24]:
```
(df1['species']=='virginica').count()
# 42 records for the Virginica species where petal_length is more than 5
```

Out[24]:
```
42
```

Find out what is the maximum median petal_length.

In [25]:
```
data.loc[(data['species']=='setosa')].median()
```

Out[25]:
```
sepal_length    5.0
sepal_width     3.4
petal_length    1.5
petal_width     0.2
dtype: float64
```

In [26]:
```
data.loc[(data['species']=='versicolor')].median()
```

Out[26]:
```
sepal_length    5.90
sepal_width     2.80
petal_length    4.35
petal_width     1.30
dtype: float64
```

In [27]:
```
data.loc[(data['species']=='virginica')].median()
```

Out[27]:
```
sepal_length    6.50
sepal_width     3.00
petal_length    5.55
petal_width     2.00
dtype: float64
```

In [ ]:
```
# 5.55 is the maximum median petal_length of virginica species
```