Most Used Pandas Commands

June 30, 2019

0.1 we will explore useful pandas commands on iris dataset

```
[68]: import pandas as pd
```

0.1.1 Basic Commands

```
[69]: # 1.Reading or Loading a datset into a variable
      dataset = pd.read_csv('Iris.csv');
```

```
[70]: #2.Displaying top 5 rows in dataset
      dataset.head()
```

[70]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

```
[71]: dataset.head(8) #Displays top 8 rows in the dataset
```

[71]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa

```
[72]: #3.Display bottom rows in the datset
      dataset.tail()
```

[72]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	\
145	146	6.7	3.0	5.2	2.3	
146	147	6.3	2.5	5.0	1.9	
147	148	6.5	3.0	5.2	2.0	
148	149	6.2	3.4	5.4	2.3	

```
149  150              5.9              3.0              5.1              1.8

          Species
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica
```

[73]: `dataset.tail(3)` *#Displays bottom 3 rows in the dataset*

[73]:
```
      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
147  148            6.5           3.0            5.2           2.0
148  149            6.2           3.4            5.4           2.3
149  150            5.9           3.0            5.1           1.8

          Species
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica
```

[74]:
```
#4.Display count of rows and columns in the dataset
dataset.shape
```

[74]: `(150, 6)`

[75]:
```
print(dataset.shape[0]) #Display rows count
print(dataset.shape[1]) #Display columns count
```

```
150
6
```

[76]:
```
#5.Display summary of the Dataset
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
Id               150 non-null int64
SepalLengthCm    150 non-null float64
SepalWidthCm     150 non-null float64
PetalLengthCm    150 non-null float64
PetalWidthCm     150 non-null float64
Species          150 non-null object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.1+ KB
```

[77]:
```
#6.Display statistical summary of all columns
dataset.describe() #statistical description of all numerical data columns
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

```python
[78]: print(dataset.describe(include = ['O']))#statistical description of categorical
      ↪columns
```

	Species
count	150
unique	3
top	Iris-setosa
freq	50

```python
[79]: print(dataset.describe(include = 'all'))#statistical description of all columns
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	\
count	150.000000	150.000000	150.000000	150.000000	150.000000	
unique	NaN	NaN	NaN	NaN	NaN	
top	NaN	NaN	NaN	NaN	NaN	
freq	NaN	NaN	NaN	NaN	NaN	
mean	75.500000	5.843333	3.054000	3.758667	1.198667	
std	43.445368	0.828066	0.433594	1.764420	0.763161	
min	1.000000	4.300000	2.000000	1.000000	0.100000	
25%	38.250000	5.100000	2.800000	1.600000	0.300000	
50%	75.500000	5.800000	3.000000	4.350000	1.300000	
75%	112.750000	6.400000	3.300000	5.100000	1.800000	
max	150.000000	7.900000	4.400000	6.900000	2.500000	

	Species
count	150
unique	3
top	Iris-setosa
freq	50
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

3

```
[80]:  #7.Display datatypes in datset
       dataset.dtypes #Datatypes of all columns in the dataset
```

```
[80]:  Id                 int64
       SepalLengthCm    float64
       SepalWidthCm     float64
       PetalLengthCm    float64
       PetalWidthCm     float64
       Species           object
       dtype: object
```

```
[81]:  dataset.Id.dtype #Datatype of particular column in the dataset
       #      (or)
       dataset['Id'].dtype
```

```
[81]:  dtype('int64')
```

```
[82]:  #8.no of uique rows in each column
       dataset.nunique() #count of unique rows in all columns
```

```
[82]:  Id               150
       SepalLengthCm     35
       SepalWidthCm      23
       PetalLengthCm     43
       PetalWidthCm      22
       Species            3
       dtype: int64
```

```
[83]:  dataset.Id.nunique() #count of unique rows of a particular column
       #      (or)
       dataset['Id'].nunique()
```

```
[83]:  150
```

```
[84]:  #9.select row with particular name or row no
       dataset.loc[:,'Id'] #select all rows in column Id
```

```
[84]:  0         1
       1         2
       2         3
       3         4
       4         5
       5         6
       6         7
       7         8
       8         9
       9        10
       10       11
       11       12
       12       13
       13       14
       14       15
```

```
15        16
16        17
17        18
18        19
19        20
20        21
21        22
22        23
23        24
24        25
25        26
26        27
27        28
28        29
29        30
         ...
120      121
121      122
122      123
123      124
124      125
125      126
126      127
127      128
128      129
129      130
130      131
131      132
132      133
133      134
134      135
135      136
136      137
137      138
138      139
139      140
140      141
141      142
142      143
143      144
144      145
145      146
146      147
147      148
148      149
149      150
Name: Id, Length: 150, dtype: int64
```

```
[85]: dataset.iloc[:,1] #select all rows in column no 1
```

```
[85]: 0      5.1
      1      4.9
      2      4.7
      3      4.6
      4      5.0
      5      5.4
      6      4.6
      7      5.0
      8      4.4
      9      4.9
      10     5.4
      11     4.8
      12     4.8
      13     4.3
      14     5.8
      15     5.7
      16     5.4
      17     5.1
      18     5.7
      19     5.1
      20     5.4
      21     5.1
      22     4.6
      23     5.1
      24     4.8
      25     5.0
      26     5.0
      27     5.2
      28     5.2
      29     4.7
            ...
      120    6.9
      121    5.6
      122    7.7
      123    6.3
      124    6.7
      125    7.2
      126    6.2
      127    6.1
      128    6.4
      129    7.2
      130    7.4
      131    7.9
      132    6.4
      133    6.3
```

```
134    6.1
135    7.7
136    6.3
137    6.4
138    6.0
139    6.9
140    6.7
141    6.9
142    5.8
143    6.8
144    6.7
145    6.7
146    6.3
147    6.5
148    6.2
149    5.9
Name: SepalLengthCm, Length: 150, dtype: float64
```

[86]:
```python
#10.find rows with duplicate data
sum(dataset.duplicated()) #count of rows with duplicated data
```

[86]: 0

[87]:
```python
dataset[dataset.duplicated()] #displays the rows with duplicate data.
```

[87]:
```
Empty DataFrame
Columns: [Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm, Species]
Index: []
```

[88]:
```python
# 11. drop a column
dataset.drop('Id',axis = 1)
```

[88]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
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
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa
10	5.4	3.7	1.5	0.2	Iris-setosa
11	4.8	3.4	1.6	0.2	Iris-setosa
12	4.8	3.0	1.4	0.1	Iris-setosa
13	4.3	3.0	1.1	0.1	Iris-setosa
14	5.8	4.0	1.2	0.2	Iris-setosa
15	5.7	4.4	1.5	0.4	Iris-setosa
16	5.4	3.9	1.3	0.4	Iris-setosa

17	5.1	3.5	1.4	0.3	Iris-setosa
18	5.7	3.8	1.7	0.3	Iris-setosa
19	5.1	3.8	1.5	0.3	Iris-setosa
20	5.4	3.4	1.7	0.2	Iris-setosa
21	5.1	3.7	1.5	0.4	Iris-setosa
22	4.6	3.6	1.0	0.2	Iris-setosa
23	5.1	3.3	1.7	0.5	Iris-setosa
24	4.8	3.4	1.9	0.2	Iris-setosa
25	5.0	3.0	1.6	0.2	Iris-setosa
26	5.0	3.4	1.6	0.4	Iris-setosa
27	5.2	3.5	1.5	0.2	Iris-setosa
28	5.2	3.4	1.4	0.2	Iris-setosa
29	4.7	3.2	1.6	0.2	Iris-setosa
..	...	...	...	...	...
120	6.9	3.2	5.7	2.3	Iris-virginica
121	5.6	2.8	4.9	2.0	Iris-virginica
122	7.7	2.8	6.7	2.0	Iris-virginica
123	6.3	2.7	4.9	1.8	Iris-virginica
124	6.7	3.3	5.7	2.1	Iris-virginica
125	7.2	3.2	6.0	1.8	Iris-virginica
126	6.2	2.8	4.8	1.8	Iris-virginica
127	6.1	3.0	4.9	1.8	Iris-virginica
128	6.4	2.8	5.6	2.1	Iris-virginica
129	7.2	3.0	5.8	1.6	Iris-virginica
130	7.4	2.8	6.1	1.9	Iris-virginica
131	7.9	3.8	6.4	2.0	Iris-virginica
132	6.4	2.8	5.6	2.2	Iris-virginica
133	6.3	2.8	5.1	1.5	Iris-virginica
134	6.1	2.6	5.6	1.4	Iris-virginica
135	7.7	3.0	6.1	2.3	Iris-virginica
136	6.3	3.4	5.6	2.4	Iris-virginica
137	6.4	3.1	5.5	1.8	Iris-virginica
138	6.0	3.0	4.8	1.8	Iris-virginica
139	6.9	3.1	5.4	2.1	Iris-virginica
140	6.7	3.1	5.6	2.4	Iris-virginica
141	6.9	3.1	5.1	2.3	Iris-virginica
142	5.8	2.7	5.1	1.9	Iris-virginica
143	6.8	3.2	5.9	2.3	Iris-virginica
144	6.7	3.3	5.7	2.5	Iris-virginica
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]
```

```
[89]: #12.count of unique values in the column
      dataset['Species'].value_counts()
```

```
[89]: Iris-setosa        50
      Iris-versicolor    50
      Iris-virginica     50
      Name: Species, dtype: int64
```

0.1.2 Advanced Commands

```
[90]: #1. query command
      dataset.query('SepalLengthCm < 5') #fetch the rows in dataset which vae␣
       ↪speallength less than 5
```

```
[90]:       Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
      1      2            4.9           3.0            1.4           0.2
      2      3            4.7           3.2            1.3           0.2
      3      4            4.6           3.1            1.5           0.2
      6      7            4.6           3.4            1.4           0.3
      8      9            4.4           2.9            1.4           0.2
      9     10            4.9           3.1            1.5           0.1
      11    12            4.8           3.4            1.6           0.2
      12    13            4.8           3.0            1.4           0.1
      13    14            4.3           3.0            1.1           0.1
      22    23            4.6           3.6            1.0           0.2
      24    25            4.8           3.4            1.9           0.2
      29    30            4.7           3.2            1.6           0.2
      30    31            4.8           3.1            1.6           0.2
      34    35            4.9           3.1            1.5           0.1
      37    38            4.9           3.1            1.5           0.1
      38    39            4.4           3.0            1.3           0.2
      41    42            4.5           2.3            1.3           0.3
      42    43            4.4           3.2            1.3           0.2
      45    46            4.8           3.0            1.4           0.3
      47    48            4.6           3.2            1.4           0.2
      57    58            4.9           2.4            3.3           1.0
      106  107            4.9           2.5            4.5           1.7

                 Species
      1      Iris-setosa
      2      Iris-setosa
      3      Iris-setosa
      6      Iris-setosa
      8      Iris-setosa
      9      Iris-setosa
      11     Iris-setosa
      12     Iris-setosa
```

```
13        Iris-setosa
22        Iris-setosa
24        Iris-setosa
29        Iris-setosa
30        Iris-setosa
34        Iris-setosa
37        Iris-setosa
38        Iris-setosa
41        Iris-setosa
42        Iris-setosa
45        Iris-setosa
47        Iris-setosa
57     Iris-versicolor
106     Iris-virginica
```

[91]:
```python
length = 1.2
dataset.query('PetalLengthCm < @length') # fetch rows with lentgh lessthan 1.2
```

[91]:
```
    Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      Species
13  14            4.3           3.0            1.1           0.1  Iris-setosa
22  23            4.6           3.6            1.0           0.2  Iris-setosa
```

[92]:
```python
#2.Get dummie variables
dataset[['Iris-setosa', 'Iris-versicolor', 'Iris-virginica']] = pd.
 ↪get_dummies(dataset['Species'])
dataset.head()
```

[92]:
```
   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      Species  \
0   1            5.1           3.5            1.4           0.2  Iris-setosa
1   2            4.9           3.0            1.4           0.2  Iris-setosa
2   3            4.7           3.2            1.3           0.2  Iris-setosa
3   4            4.6           3.1            1.5           0.2  Iris-setosa
4   5            5.0           3.6            1.4           0.2  Iris-setosa

   Iris-setosa  Iris-versicolor  Iris-virginica
0            1                0               0
1            1                0               0
2            1                0               0
3            1                0               0
4            1                0               0
```

[93]:
```python
#3.group by
dataset.groupby('Species').count()
```

[93]:
```
                 Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
Species
Iris-setosa      50             50            50             50            50
Iris-versicolor  50             50            50             50            50
Iris-virginica   50             50            50             50            50
```

```
                    Iris-setosa  Iris-versicolor  Iris-virginica
    Species
    Iris-setosa                 50               50              50
    Iris-versicolor             50               50              50
    Iris-virginica              50               50              50
```

[94]:
```python
#4. order the datset
dataset.sort_values(by = 'Species', inplace = True)
dataset.head()
```

[94]:
```
        Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      Species  \
    0    1            5.1           3.5            1.4           0.2  Iris-setosa
    27  28            5.2           3.5            1.5           0.2  Iris-setosa
    28  29            5.2           3.4            1.4           0.2  Iris-setosa
    29  30            4.7           3.2            1.6           0.2  Iris-setosa
    30  31            4.8           3.1            1.6           0.2  Iris-setosa

        Iris-setosa  Iris-versicolor  Iris-virginica
    0             1                0               0
    27            1                0               0
    28            1                0               0
    29            1                0               0
    30            1                0               0
```

[95]:
```python
#5.rename the column
dataset.rename(columns = {"Id":"id","Species":"species"},inplace = True)
dataset.head()
```

[95]:
```
        id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      species  \
    0    1            5.1           3.5            1.4           0.2  Iris-setosa
    27  28            5.2           3.5            1.5           0.2  Iris-setosa
    28  29            5.2           3.4            1.4           0.2  Iris-setosa
    29  30            4.7           3.2            1.6           0.2  Iris-setosa
    30  31            4.8           3.1            1.6           0.2  Iris-setosa

        Iris-setosa  Iris-versicolor  Iris-virginica
    0             1                0               0
    27            1                0               0
    28            1                0               0
    29            1                0               0
    30            1                0               0
```

[96]:
```python
dataset.rename(str.lower,axis ='columns', inplace = True)
dataset.head()
```

[96]:
```
        id  sepallengthcm  sepalwidthcm  petallengthcm  petalwidthcm      species  \
    0    1            5.1           3.5            1.4           0.2  Iris-setosa
    27  28            5.2           3.5            1.5           0.2  Iris-setosa
    28  29            5.2           3.4            1.4           0.2  Iris-setosa
    29  30            4.7           3.2            1.6           0.2  Iris-setosa
```

```
30  31                4.8            3.1            1.6          0.2  Iris-setosa
```

```
      iris-setosa   iris-versicolor   iris-virginica
0            1                0                0
27           1                0                0
28           1                0                0
29           1                0                0
30           1                0                0
```

[97]:
```python
#6. merge datsets
df1 = dataset[['id','sepallengthcm','sepalwidthcm']]
print('df1')
print(df1.columns)

df2 = dataset[['id','petallengthcm','petalwidthcm','species']]
print('df2')
print(df2.columns)

new_data = pd.merge(df1,df2,on = 'id')
print(new_data.head())
print('')
print(new_data.shape)
```

```
df1
Index(['id', 'sepallengthcm', 'sepalwidthcm'], dtype='object')
df2
Index(['id', 'petallengthcm', 'petalwidthcm', 'species'], dtype='object')
    id  sepallengthcm  sepalwidthcm  petallengthcm  petalwidthcm      species
0    1            5.1           3.5            1.4           0.2  Iris-setosa
1   28            5.2           3.5            1.5           0.2  Iris-setosa
2   29            5.2           3.4            1.4           0.2  Iris-setosa
3   30            4.7           3.2            1.6           0.2  Iris-setosa
4   31            4.8           3.1            1.6           0.2  Iris-setosa

(150, 6)
```

[98]:
```python
#7. concat datsets
df1 = dataset.iloc[0:75,:]
print('df1')
print(df1.head(3))
print('-----------------------------------------------------------------------

df2 = dataset.iloc[75:150,:]
print('df2')
print(df2.head(3))
print('-----------------------------------------------------------------------
```

```
df1
```

```
       id  sepallengthcm  sepalwidthcm  petallengthcm  petalwidthcm       species  \
0       1            5.1           3.5            1.4           0.2  Iris-setosa
27     28            5.2           3.5            1.5           0.2  Iris-setosa
28     29            5.2           3.4            1.4           0.2  Iris-setosa

       iris-setosa  iris-versicolor  iris-virginica
0               1                0               0
27              1                0               0
28              1                0               0
--------------------------------------------------------------------------------
-------
df2
       id  sepallengthcm  sepalwidthcm  petallengthcm  petalwidthcm  \
74     75            6.4           2.9            4.3           1.3
73     74            6.1           2.8            4.7           1.2
51     52            6.4           3.2            4.5           1.5

              species  iris-setosa  iris-versicolor  iris-virginica
74  Iris-versicolor            0                1               0
73  Iris-versicolor            0                1               0
51  Iris-versicolor            0                1               0
--------------------------------------------------------------------------------
---------
```

[99]:
```
new_data = pd.concat([df1,df2])

print(new_data.shape)
print('')
new_data.head(3)
```

```
(150, 9)
```

[99]:
```
       id  sepallengthcm  sepalwidthcm  petallengthcm  petalwidthcm       species  \
0       1            5.1           3.5            1.4           0.2  Iris-setosa
27     28            5.2           3.5            1.5           0.2  Iris-setosa
28     29            5.2           3.4            1.4           0.2  Iris-setosa

       iris-setosa  iris-versicolor  iris-virginica
0               1                0               0
27              1                0               0
28              1                0               0
```
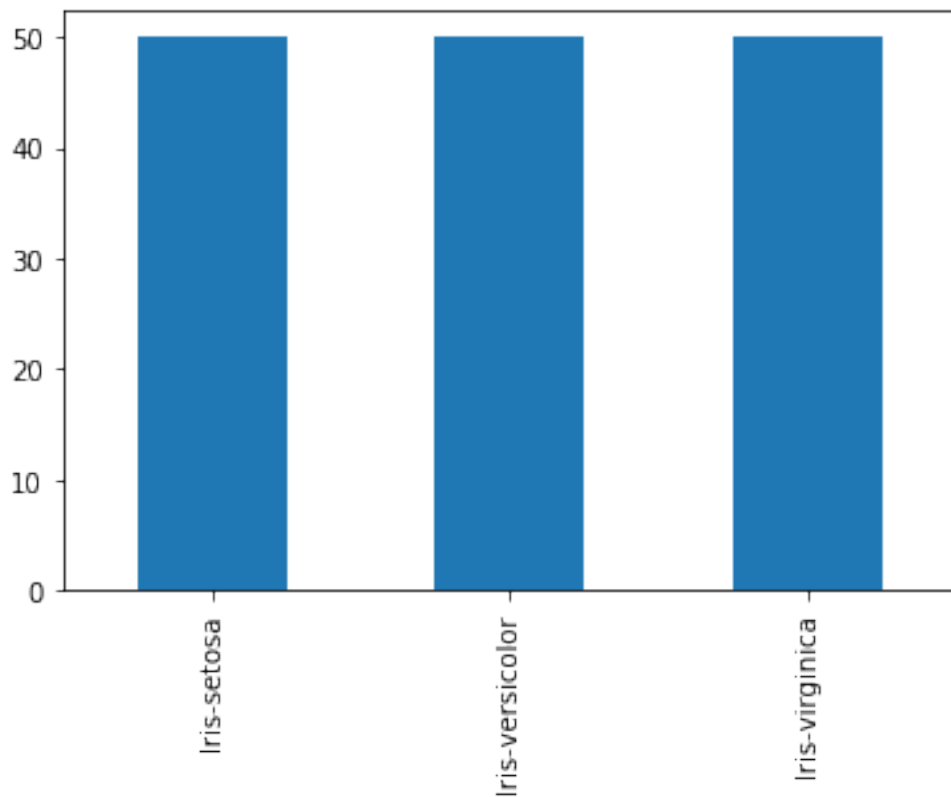
0.1.3 Plotting with pandas

```
[100]: #1. Scatter plot
       dataset.plot.scatter('sepallengthcm','petallengthcm');
```
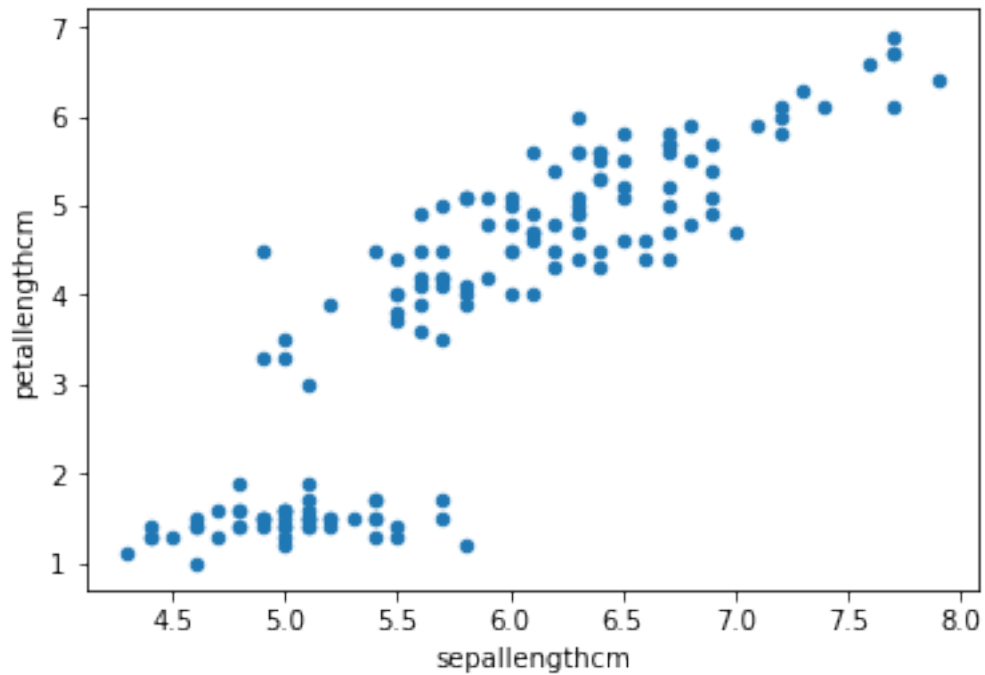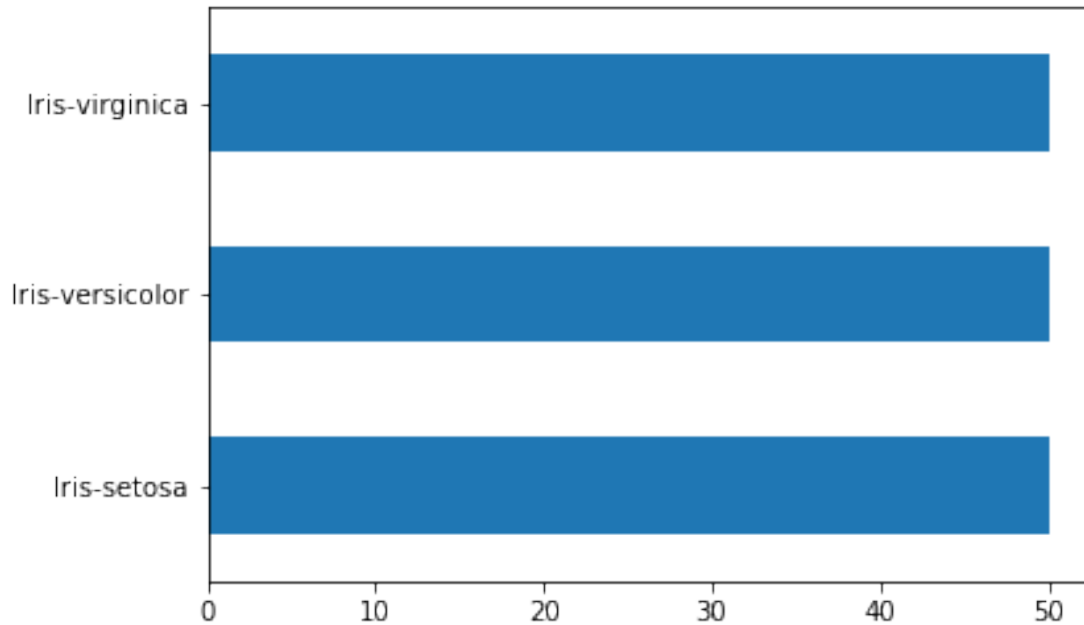


```
[101]: #2. Histogram plot
       dataset['sepalwidthcm'].plot.hist();
```
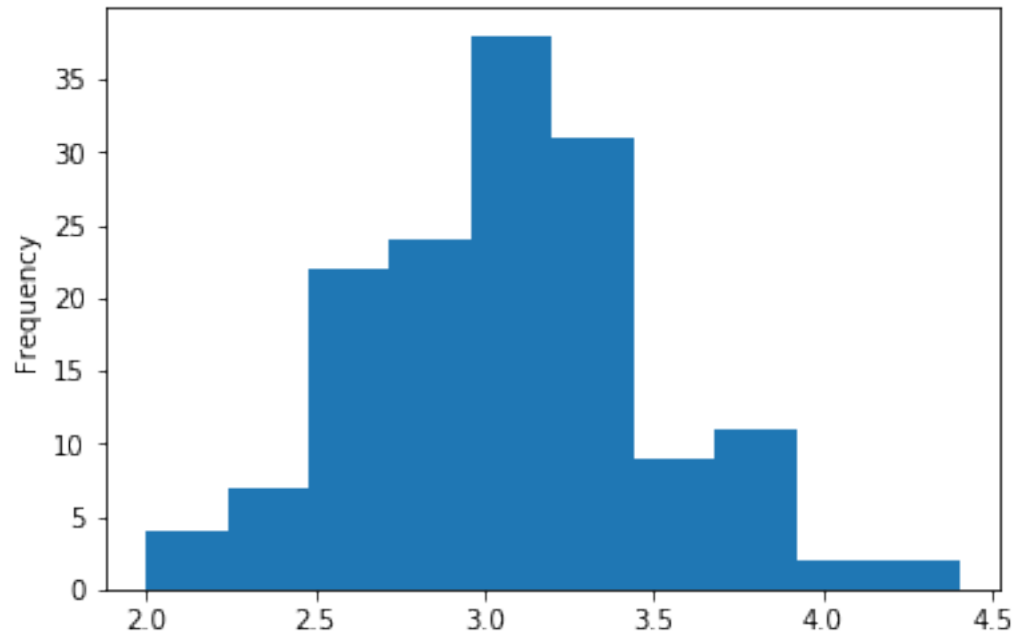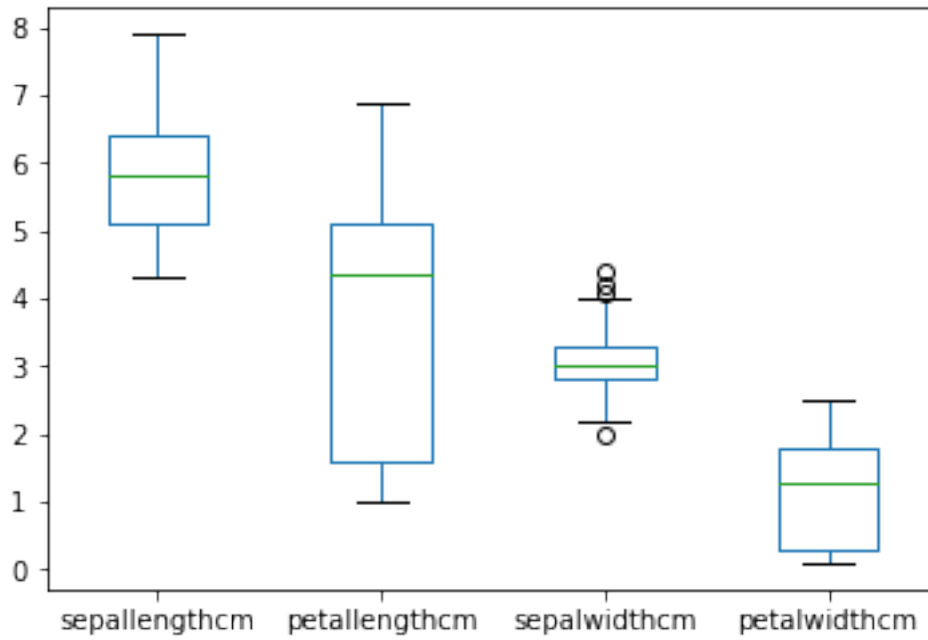
[102]:
```
#3. Bar plot
dataset['species'].value_counts().plot.bar();
```
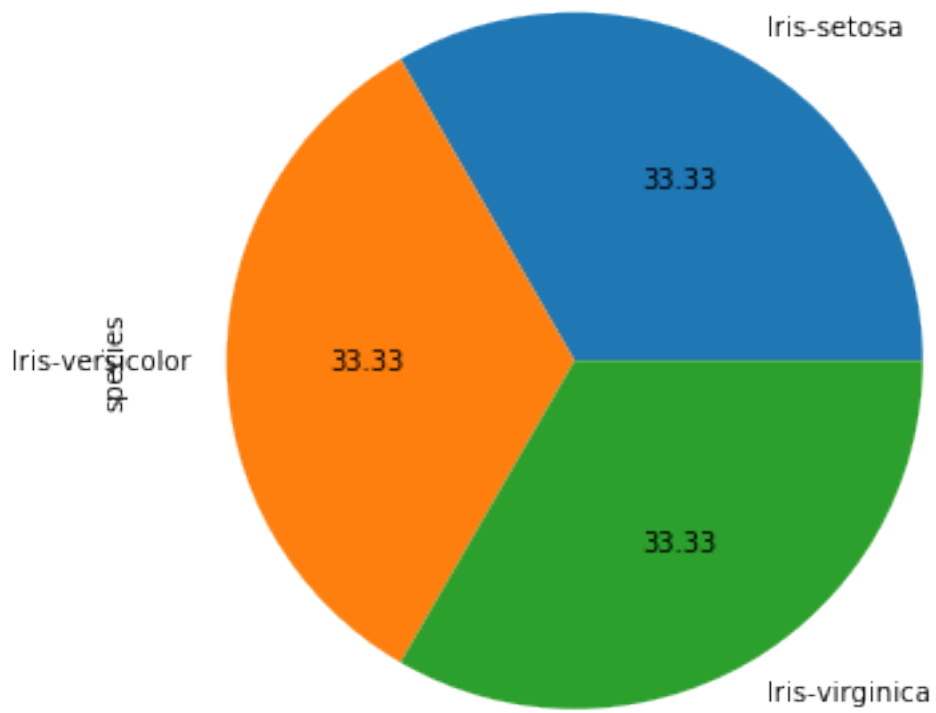
```
[103]: #4. Horizontal bar plot
        dataset['species'].value_counts().plot.barh();
```



```
[104]: #5. box plots
        dataset[['sepallengthcm','petallengthcm','sepalwidthcm','petalwidthcm']].plot.
         ↪box();
```

[105]:
```
#6 pie plots
dataset['species'].value_counts().plot.pie(autopct='%.2f',figsize=(6,6));
```

find all kinds of pandas visualiations in the link