

we will explore useful pandas commands on iris dataset

In [68]: `import pandas as pd`

Basic Commands

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

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

	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

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

	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

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

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.3	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

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

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
147	148	6.5	3.0	5.2	2.3	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

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

Out [74]: `(150, 6)`

In [75]: `print(dataset.shape[0]) #display row count  
print(dataset.shape[1]) #display column count`

150  
6

In [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

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

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	
count	150	150.000000	150.000000	150.000000	150.000000	
mean	75.000000	5.843333	3.854000	3.758667	1.198667	
std	43.440368	0.828066	0.433594	1.764420	0.763161	
min	1	4.3	2.5	1.3	0.2	Iris-setosa
max	150	7.9	4.4	6.7	2.3	Iris-virginica

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

Species  
count 150  
unique 3  
top Iris-setosa  
freq 50

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

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
count	150	150.000000	150.000000	150.000000	150.000000	150.000000
unique	NaN	NaN	NaN	NaN	NaN	NaN
top	NaN	NaN	NaN	NaN	NaN	NaN
freq	NaN	NaN	NaN	NaN	NaN	NaN
mean	75.000000	5.843333	3.854000	3.758667	1.198667	1.198667
std	43.440368	0.828066	0.433594	1.764420	0.763161	0.763161
min	1	4.3	2.5	1.3	0.2	Iris-setosa
max	150	7.9	4.4	6.7	2.3	Iris-virginica

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

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

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

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

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

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

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

Out [83]: `150`

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

Out [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`

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

Out [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 5.7  
18 5.7  
19 5.1  
20 5.4  
21 5.1  
22 4.6  
23 5.1  
24 6.8  
25 5.0  
26 5.0  
27 5.2  
28 5.2  
29 4.7  
...  
120 6.9  
121 6.9  
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`

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

Out [86]: `0`

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

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
--	----	---------------	--------------	---------------	--------------	---------

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

	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	2.8	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	5.8	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	1.9	Iris-virginica
127	6.1	3.0	4.8	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	5.8	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

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

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

Advanced Commands

In [90]: `#1. query command  
dataset.query('SepalLengthCm < 5') #fetch the rows in dataset which var SepalLength less than 5`

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
--	----	---------------	--------------	---------------	--------------	---------

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
6	7	4.6	3.4	1.4	0.3	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	4.3	3.0	1.1	0.1	Iris-setosa
22	23	4.6	3.6	1.0	0.2	Iris-setosa
24	25	4.8	3.4	1.9	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
34	35	4.9	3.1	1.5	0.1	Iris-setosa
37	38	4.9	3.1	1.5	0.1	Iris-setosa
38	39	4.4	3.0	1.3	0.2	Iris-setosa
41	42	4.5	2.3	1.3	0.3	Iris-setosa
42	43	4.4	3.2	1.3	0.2	Iris-setosa
45	46	4.8	3.0	1.4	0.3	Iris-setosa
47	48	4.6	3.2	1.4	0.2	Iris-setosa
147	148	4.8	2.4	5.3	2.0	Iris-virginica
148	149	4.9	2.5	4.5	1.7	Iris-virginica

In [91]: `length = 1.2  
dataset.query('PetalLengthCm < @length') # fetch rows with length less than 1.2`

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
--	----	---------------	--------------	---------------	--------------	---------

Out [91]: `13 14 4.3 3.0 1.1 0.1 Iris-setosa  
22 23 4.6 3.6 1.0 0.2 Iris-setosa`

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

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Iris-setosa	Iris-versicolor	Iris-virginica
--	----	---------------	--------------	---------------	--------------	---------	-------------	-----------------	----------------

0	1	5.1	3.5	1.4	0.2	Iris-setosa	1	0	0
1	2	4.9	3.0	1.4	0.2	Iris-setosa	1	0	0
2	3	4.7	3.2	1.3	0.2	Iris-setosa	1	0	0
3	4	4.6	3.1	1.5	0.2	Iris-setosa	1	0	0
4	5	5.0	3.6	1.4	0.2	Iris-setosa	1	0	0

In [93]: `#2. group by  
dataset.groupby('Species').count()`

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Iris-setosa	Iris-versicolor	Iris-virginica
--	----	---------------	--------------	---------------	--------------	-------------	-----------------	----------------

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

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

	id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Iris-setosa	Iris-versicolor	Iris-virginica
--	----	---------------	--------------	---------------	--------------	---------	-------------	-----------------	----------------

```
In [97]: #6. merge datasets
df1 = dataset[['id', 'sepal.length', 'sepal.width']]
print(df1)
```