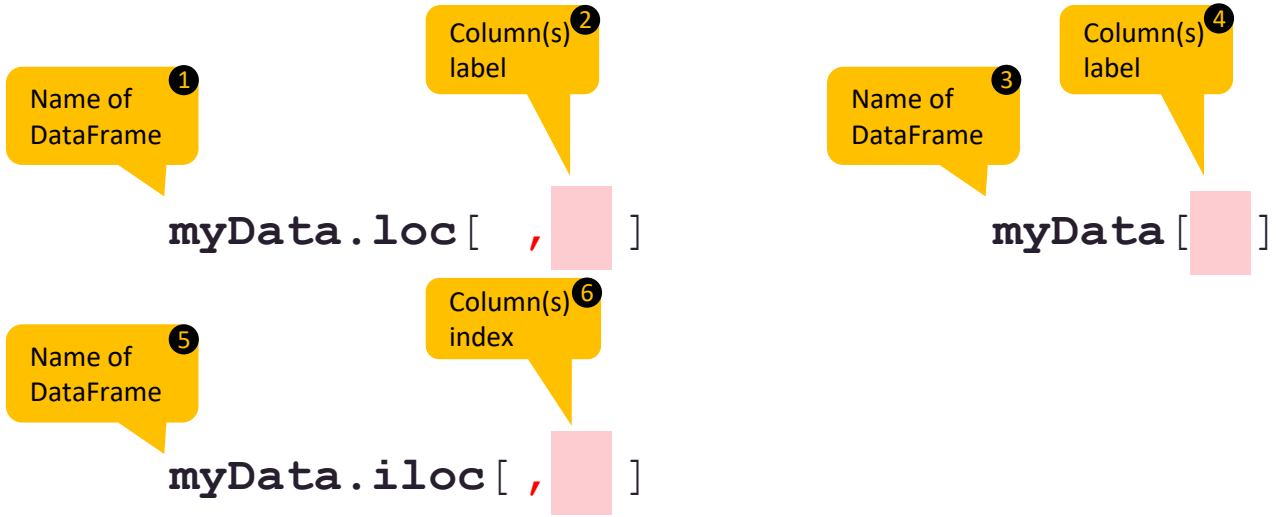
Selecting columns

There are multiple ways of selecting columns

1. Select a single column
2. Select multiple columns
3. Combine operations to select by rows and columns

Enter your selection commands in the column placeholder

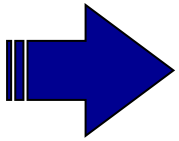
Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...



Select a single column by column name / number (1/2)

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Select column `TransDate`



TransDate
15.11.2005
29.08.2008
19.10.2007
14.11.2005
12.06.2007
...

1 Returns a DataFrame

2 DataFrame name

3 All rows

4 Column name

```
myData.loc[ : , "TransDate"]
myData.iloc[ : , 1]
myData ["TransDate"]
```

5 Select the second column (corresponding index=1)

Select a single column by column name / number (2/2)

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Select column `TransDate` and return it as numpy array

```
array(['2005-11-
15T00:00:00.000000000',
'2008-08-
29T00:00:00.000000000',
'2007-10-
19T00:00:00.000000000', ...,
dtype='datetime64[ns]')
```

1 Returns a numpy array

2 Turn the output into array

```
myData.loc[ : , "TransDate"].values
myData.iloc[ : , 1].values
myData ["TransDate"].values
```

Select a single column by column name / number (2/2)

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Select column `TransDate` and return it as numpy array

```
array(['2005-11-
15T00:00:00.000000000',
'2008-08-
29T00:00:00.000000000',
'2007-10-
19T00:00:00.000000000', ...,
dtype='datetime64[ns]')
```

1 Returns a numpy array

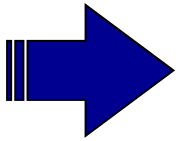
2 Turn the output into array

```
myData.loc[ : , "TransDate"].values

myData.iloc[ : , 1].values

myData ["TransDate"].values
```

Select multiple columns by column name / number

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Select columns `Customer`, `TransDate`, and `PurchAmount`



Customer	TransDate	PurchAmount
149332	15.11.2005	199.95
172951	29.08.2008	199.95
120621	19.10.2007	99.95
149236	14.11.2005	39.95
149236	12.06.2007	79.95
...	...	...

1 List of column vectors

2 Column names

```
myData.loc[ : , ["Customer", "TransDate", "PurchAmount"]]
```

```
myData.iloc[ : , [0,1,3]]
```

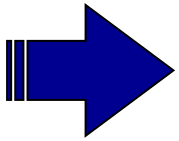
3 Column indexes

```
myData[["Customer", "TransDate", "PurchAmount"]]
```

Select multiple columns by column name / number

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Select columns `Customer`, `TransDate`, and `PurchAmount`



Customer	TransDate	PurchAmount
149332	15.11.2005	199.95
172951	29.08.2008	199.95
120621	19.10.2007	99.95
149236	14.11.2005	39.95
149236	12.06.2007	79.95
...	...	...

1 List of column vectors

2 Column names

```
myData.loc[ : , ["Customer", "TransDate", "PurchAmount"]]
```

```
myData.iloc[ : , [0,1,3]]
```

3 Column indexes

```
myData[["Customer", "TransDate", "PurchAmount"]]
```

Python Basics: Find out the column names

Get all column names from your DataFrame:

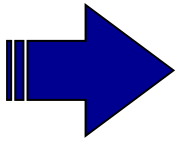
```
> myData.columns.values
```

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Combine operations to select by rows and columns

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Select column `TransDate` and `Cost` for entries where `PurchAmount` > 100



TransDate	Cost
15.11.2005	107.00
29.08.2008	108.00
...	...

1 Filtering condition for rows

2 Select columns

```
myData.loc[ myData["PurchAmount"]>100 , ["TransDate","Cost"]]
```

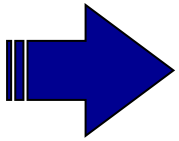
```
myData[["TransDate","Cost"]].loc[ myData["PurchAmount"]>100 ,]
```

3 Select columns first

Combine operations to select by rows and columns

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Select column `TransDate` and `Cost` for entries where `PurchAmount` > 100



TransDate	Cost
15.11.2005	107.00
29.08.2008	108.00
...	...

1 Filtering condition for rows

2 Select columns

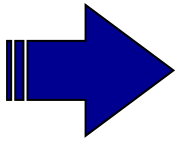
```
myData.loc[ myData["PurchAmount"]>100 , ["TransDate","Cost"]]
myData[["TransDate","Cost"]].loc[ myData["PurchAmount"]>100 ,]
```

3 Select columns first

Combine operations to select by rows and columns

Customer	TransDate	Quantity	PurchAmount	Cost	TransID
149332	15.11.2005	1	199.95	107.00	127998739
172951	29.08.2008	1	199.95	108.00	128888288
120621	19.10.2007	1	99.95	49.00	125375247
149236	14.11.2005	1	39.95	18.95	127996226
149236	12.06.2007	1	79.95	35.00	128670302
...	...	...	...	...	...

Select column `TransDate` and `Cost` for entries where `PurchAmount` > 100



TransDate	Cost
15.11.2005	107.00
29.08.2008	108.00
...	...

Filtering condition for rows (1)

Select columns (2)

```
myData.loc[ myData["PurchAmount"]>100 , ["TransDate","Cost"]]
```

```
myData[["TransDate","Cost"]].loc[ myData["PurchAmount"]>100 ,]
```

Select columns first (3)

Selecting columns