

Working with multiple sheets

PYTHON FOR SPREADSHEET USERS



Chris Cardillo
Data Scientist at DataCamp

Previously

```
# Import package
import pandas as pd

# Read file
fruit = pd.read_excel('fruit.xlsx')

# Print to console
print(fruit)
```

	name	color	price_usd
0	Apple	red	0.88
1	Banana	yellow	0.23
2	Orange	orange	0.68
3	Watermelon	green	3.98
4	Plum	purple	0.96
5	Blueberries	blue	5.16
6	Dragonfruit	pink	5.27
7	Kiwi	brown	1.12

Previously

	A	B	C	D
1	name	color	price_usd	
2	Apple	red	0.88	
3	Banana	yellow	0.23	
4	Orange	orange	0.68	
5	Watermelon	green	3.98	
6	Plum	purple	0.96	
7	Blueberries	blue	5.16	
8	Dragonfruit	pink	5.27	
9	Kiwi	brown	1.12	
10				

Two Tabs

PRICE TAB

	A	B	C
1	name	price_usd	
2	Apple	0.88	
3	Banana	0.23	
4	Orange	0.68	
5	Watermelon	3.98	
6	Plum	0.96	
7	Blueberries	5.16	
8	Dragonfruit	5.27	
9	Kiwi	1.12	
10			
11			
12			

COLOR TAB

	A	B	C
1	name	color	
2	Apple	red	
3	Banana	yellow	
4	Orange	orange	
5	Watermelon	green	
6	Plum	purple	
7	Blueberries	blue	
8	Dragonfruit	pink	
9	Kiwi	brown	
10			
11			
12			

pd.ExcelFile()

```
# Import package
import pandas as pd

# Read workbook
fruit_workbook = pd.ExcelFile('fruit_tabs.xlsx')
```

pd.ExcelFile()

```
# Import package
import pandas as pd

# Read workbook
fruit_workbook = pd.ExcelFile('fruit_tabs.xlsx')

# Print to console
print(fruit_workbook)
```

```
<pandas.io.excel.ExcelFile object at 0x7fe9d32e9dd8>
```

Attributes

- Live within objects, just like methods
- Accessed with the dot
- No parentheses required

.sheet_names attribute

```
# Read workbook  
fruit_workbook = pd.ExcelFile('fruit_tabs.xlsx')
```


.sheet_names attribute

```
# Read workbook
fruit_workbook = pd.ExcelFile('fruit_tabs.xlsx')

# Get sheet names
fruit_sheet_names = fruit_workbook.sheet_names
```

.sheet_names attribute

```
# Read workbook
fruit_workbook = pd.ExcelFile('fruit_tabs.xlsx')

# Get sheet names
fruit_sheet_names = fruit_workbook.sheet_names

# Print sheet names to console
print(fruit_sheet_names)
```

```
['price', 'color']
```

Attributes vs. methods

ATTRIBUTES

- `object.attribute`
- `workbook.sheet_names`
- Tell us something

METHODS

- `object.method()`
- `workbook.parse()`
- Do something for us

Attributes vs. methods

ATTRIBUTES

- `object.attribute`
- `workbook.sheet_names`
- Tell us something
- Always attached to an object!

METHODS

- `object.method()`
- `workbook.parse()`
- Do something for us
- Always attached to an object!

Attributes vs. methods vs. functions

ATTRIBUTES

- `object.attribute`
- `workbook.sheet_names`
- Tell us something
- Always attached to an object!

METHODS

- `object.method()`
- `workbook.parse()`
- Do something for us
- Always attached to an object!

FUNCTIONS

- `function()` or `package.function()`
- `pd.ExcelFile()`

.parse() method

```
# Read workbook
fruit_workbook = pd.ExcelFile('fruit_tabs.xlsx')

# Parse price tab
fruit_prices = fruit_workbook.parse('price')
```

.parse() method

```
# Read workbook
fruit_workbook = pd.ExcelFile('fruit_tabs.xlsx')

# Parse price tab
fruit_prices = fruit_workbook.parse('price')

# Print fruit prices
print(fruit_prices)
```

	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5	Blueberries	5.16

Recap

- `pd.ExcelFile()` function
- `workbook.sheet_names` attribute
- `workbook.parse()` method

Your Turn!

PYTHON FOR SPREADSHEET USERS

Preparing to put tables together

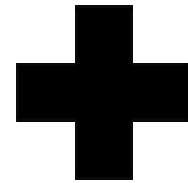
PYTHON FOR SPREADSHEET USERS



Chris Cardillo
Data Scientist at DataCamp

Table Merging

	name	color
0	Apple	red
1	Banana	yellow
2	Orange	orange
3	Watermelon	green
4	Plum	purple
5	Blueberries	blue
6	Dragonfruit	pink
7	Kiwi	brown

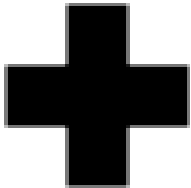


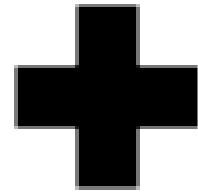
	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5	Blueberries	5.16
6	Dragonfruit	5.27
7	Kiwi	1.12



	name	color	price_usd
0	Apple	red	0.88
1	Banana	yellow	0.23
2	Orange	orange	0.68
3	Watermelon	green	3.98
4	Plum	purple	0.96
5	Blueberries	blue	5.16
6	Dragonfruit	pink	5.27
7	Kiwi	brown	1.12

Table Merging

	name	color		name	price_usd		name	color	price_usd	
0	Apple	red		0	Apple	0.88	0	Apple	red	0.88
1	Banana	yellow		1	Banana	0.23	1	Banana	yellow	0.23
2	Orange	orange		2	Orange	0.68	2	Orange	orange	0.68
3	Watermelon	green		3	Watermelon	3.98	3	Watermelon	green	3.98
4	Plum	purple		4	Plum	0.96	4	Plum	purple	0.96
5	Blueberries	blue		5	Blueberries	5.16	5	Blueberries	blue	5.16
6	Dragonfruit	pink		6	Dragonfruit	5.27	6	Dragonfruit	pink	5.27
7	Kiwi	brown		7	Kiwi	1.12	7	Kiwi	brown	1.12



Merging: The VLOOKUP of Data Science

fx | =VLOOKUP(A2,E2:F9,2,0)

	A	B	C	D	E	F
1	name	color	price_usd		name	price_usd
2	Apple	red	0.88		Apple	0.88
3	Banana	yellow			Banana	0.23
4	Orange	orange			Orange	0.68
5	Watermelon	green			Watermelon	3.98
6	Plum	purple			Plum	0.96
7	Blueberries	blue			Blueberries	5.16
8	Dragonfruit	pink			Dragonfruit	5.27
9	Kiwi	brown			Kiwi	1.12

Merging: The VLOOKUP of Data Science

<i>fx</i>	=VLOOKUP(A2,E2:F9,2,0)					
	A	B	C	D	E	F
1	name	color	price_usd		name	price_usd
2	Apple	red	0.88		Apple	0.88
3	Banana	yellow			Banana	0.23
4	Orange	orange			Orange	0.68
5	Watermelon	green			Watermelon	3.98
6	Plum	purple			Plum	0.96
7	Blueberries	blue			Blueberries	5.16
8	Dragonfruit	pink			Dragonfruit	5.27
9	Kiwi	brown			Kiwi	1.12

Merging: The VLOOKUP of Data Science

<i>fx</i>	<code>=VLOOKUP(A2,E2:F9,2,0)</code>					
	A	B	C	D	E	F
1	name	color	price_usd		name	price_usd
2	Apple	red	0.88		Apple	0.88
3	Banana	yellow			Banana	0.23
4	Orange	orange			Orange	0.68
5	Watermelon	green			Watermelon	3.98
6	Plum	purple			Plum	0.96
7	Blueberries	blue			Blueberries	5.16
8	Dragonfruit	pink			Dragonfruit	5.27
9	Kiwi	brown			Kiwi	1.12

Merging in Python

- Python is case-sensitive!
 - `'Apple' != 'apple'`

Merging in Python

- Python is case-sensitive!
 - `'Apple' != 'apple'`
- Python is exact with whitespace!
 - `'Apple' != ' Apple '`

Merging in Python

- Python is case-sensitive!
 - `'Apple' != 'apple'`
- Python is exact!
 - `'Apple' != ' Apple '`
- Python joins whole tables!
 - $AB + ACD = ABCD$, even if we only want ABC
 - So we have to drop column D.

Handling case-sensitivity

FRUIT_PRICE - UPPERCASE NAME

	name	price_usd
0	APPLE	0.88
1	BANANA	0.23
2	ORANGE	0.68
3	WATERMELON	3.98
4	PLUM	0.96
5	BLUEBERRIES	5.16
6	DRAGONFRUIT	5.27
7	KIWI	1.12

FRUIT_COLOR - TITLE CASE NAME

	name	color
0	Apple	red
1	Banana	yellow
2	Orange	orange
3	Watermelon	green
4	Plum	purple
5	Blueberries	blue
6	Dragonfruit	pink

.str.title() method

```
fruit_price['name'].str.title()
```

.str.title() method

```
fruit_price['name'] = fruit_price['name'].str.title()
```

```
print(fruit_price)
```

	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5	Strawberry	1.45

.str.upper() method

```
fruit_price['name'] = fruit_price['name'].str.upper()
```

```
print(fruit_price)
```

	name	price_usd
0	APPLE	0.88
1	BANANA	0.23
2	ORANGE	0.68
3	WATERMELON	3.98
4	PLUM	0.96
5	BLUEBERRY	5.45

.str.lower() method

```
fruit_price['name'] = fruit_price['name'].str.lower()
```

```
print(fruit_price)
```

	name	price_usd
0	apple	0.88
1	banana	0.23
2	orange	0.68
3	watermelon	3.98
4	plum	0.96

.str.title() method

```
fruit_price['name'] = fruit_price['name'].str.title()
```

```
print(fruit_price)
```

	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5

Handling whitespace with .str.strip()

```
fruit_price['name'] = fruit_price['name'].str.strip()

print(fruit_price)
```

	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5		

Selecting and dropping columns

```
      name  price_usd  unneeded
0     Apple         0.88    useless
1   Banana         0.23    useless
2   Orange         0.68    useless
3 Watermelon        3.98    useless
4      Plum         0.96    useless
5  Dishwasher        1.1    useless
```

Selecting columns

SELECT COLUMNS

```
fruit_price[['name', 'price_usd']]
```

	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5	Blueberries	5.16

Selecting columns

SELECT COLUMNS

```
fruit_price = fruit_price[['name', 'price_usd']]
```

	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5	Blueberries	5.16

Dropping Columns with .drop()

DROP COLUMNS

```
fruit_price.drop('unneeded', axis=1)
```

	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5	Blueberries	5.16

Dropping Columns

DROP COLUMNS

```
fruit_price = fruit_price.drop('unneeded', axis=1)
```

	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5	Blueberries	5.16

Your turn!

PYTHON FOR SPREADSHEET USERS

Merging: The VLOOKUP of Python

PYTHON FOR SPREADSHEET USERS



Chris Cardillo
Data Scientist at DataCamp

VLOOKUP

fx | =VLOOKUP(A5, \$E\$2:\$F\$9, 2, 0)

	A	B	C	D	E	F
1	name	color	price_usd		name	price_usd
2	Apple	red	0.88		Apple	0.88
3	Banana	yellow	0.23		Banana	0.23
4	Orange	orange	0.68		Orange	0.68
5	Watermelon	green	3.98		Watermelon	3.98
6	Plum	purple	0.96		Plum	0.96
7	Blueberries	blue	5.16		Blueberries	5.16
8	Dragonfruit	pink	5.27		Dragonfruit	5.27
9	Kiwi	brown	1.12		Kiwi	1.12

VLOOKUP - the key column

fx | =VLOOKUP(A5, \$E\$2:\$F\$9, 2, 0)

	A	B	C	D	E	F
1	name	color	price_usd		name	price_usd
2	Apple	red	0.88		Apple	0.88
3	Banana	yellow	0.23		Banana	0.23
4	Orange	orange	0.68		Orange	0.68
5	Watermelon	green	3.98		Watermelon	3.98
6	Plum	purple	0.96		Plum	0.96
7	Blueberries	blue	5.16		Blueberries	5.16
8	Dragonfruit	pink	5.27		Dragonfruit	5.27
9	Kiwi	brown	1.12		Kiwi	1.12

VLOOKUP - A left join

<i>fx</i>	<code>=VLOOKUP(A5, \$E\$2:\$F\$9, 2, 0)</code>					
	A	B	C	D	E	F
1	name	color	price_usd		name	price_usd
2	Apple	red	0.88	←	Apple	0.88
3	Banana	yellow	0.23		Banana	0.23
4	Orange	orange	0.68		Orange	0.68
5	Watermelon	green	3.98	←	Watermelon	3.98
6	Plum	purple	0.96		Plum	0.96
7	Blueberries	blue	5.16		Blueberries	5.16
8	Dragonfruit	pink	5.27	←	Dragonfruit	5.27
9	Kiwi	brown	1.12		Kiwi	1.12

VLOOKUP - #N/A

fx | =VLOOKUP(A5, \$E\$2:\$F\$9, 2, 0)

	A	B	C	D	E	F
1	name	color	price_usd		name	price_usd
2	Apple	red	0.88	←	Apple	0.88
3	Banana	yellow	0.23		Banana	0.23
4	Orange	orange	0.68		Orange	0.68
5	Watermelon	green	3.98	←	Watermelon	3.98
6	Plum	purple	0.96		Plum	0.96
7	Blueberries	blue	5.16		Blueberries	5.16
8	Dragonfruit	pink	5.27	←	Dragonfruit	5.27
9	Kiwi	brown	#N/A	↙		

The .merge() method

```
- -
      name    color
0      Apple      red
1     Banana  yellow
2      Orange  orange
3  Watermelon  green
4         Plum  purple
5  Blueberries   blue
6  Dragonfruit  pink
7         Kiwi   brown

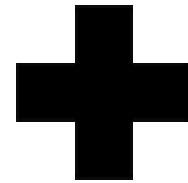
      name    price_usd
0      Apple      0.88
1     Banana      0.23
2      Orange      0.68
3  Watermelon      3.98
4         Plum      0.96
5  Blueberries      5.16
6  Dragonfruit      5.27
7         Kiwi      1.12
```

fruit_prices

The .merge() method

```
fruit_colors.merge(fruit_prices, on='name', how='left')
```

	name	color
0	Apple	red
1	Banana	yellow
2	Orange	orange
3	Watermelon	green
4	Plum	purple
5	Blueberries	blue
6	Dragonfruit	pink
7	Kiwi	brown



	name	price_usd
0	Apple	0.88
1	Banana	0.23
2	Orange	0.68
3	Watermelon	3.98
4	Plum	0.96
5	Blueberries	5.16
6	Dragonfruit	5.27
7	Kiwi	1.12



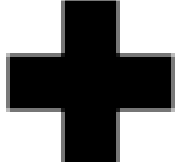

	name	color	price_usd
0	Apple	red	0.88
1	Banana	yellow	0.23
2	Orange	orange	0.68
3	Watermelon	green	3.98
4	Plum	purple	0.96
5	Blueberries	blue	5.16
6	Dragonfruit	pink	5.27
7	Kiwi	brown	1.12

Key columns without the same name

	name	color		fruit_type	price_usd
0	Apple	red	0	Apple	0.88
1	Banana	yellow	1	Banana	0.23
2	Orange	orange	2	Orange	0.68
3	Watermelon	green	3	Watermelon	3.98
4	Plum	purple	4	Plum	0.96
5	Blueberries	blue	5	Blueberries	5.16
6	Dragonfruit	pink	6	Dragonfruit	5.27
7	Kiwi	brown	7	Kiwi	1.12

`left_on` and `right_on` in .merge()

```
fruit_colors.merge(fruit_prices,  
                   left_on='name',  
                   right_on='fruit_type',  
                   how='left')
```

	name	color		fruit_type	price_usd		name	color	fruit_type	price_usd		
0	Apple	red		0	Apple	0.88	0	Apple	red	Apple	0.88	
1	Banana	yellow		1	Banana	0.23	1	Banana	yellow	Banana	0.23	
2	Orange	orange		2	Orange	0.68		2	Orange	orange	Orange	0.68
3	Watermelon	green		3	Watermelon	3.98	3	Watermelon	green	Watermelon	3.98	
4	Plum	purple		4	Plum	0.96	4	Plum	purple	Plum	0.96	
5	Blueberries	blue		5	Blueberries	5.16	5	Blueberries	blue	Blueberries	5.16	
6	Dragonfruit	pink		6	Dragonfruit	5.27	6	Dragonfruit	pink	Dragonfruit	5.27	
7	Kiwi	brown	7	Kiwi	1.12	7	Kiwi	brown	Kiwi	1.12		

Your turn!

PYTHON FOR SPREADSHEET USERS