

Welcome to Python!

PYTHON FOR SPREADSHEET USERS

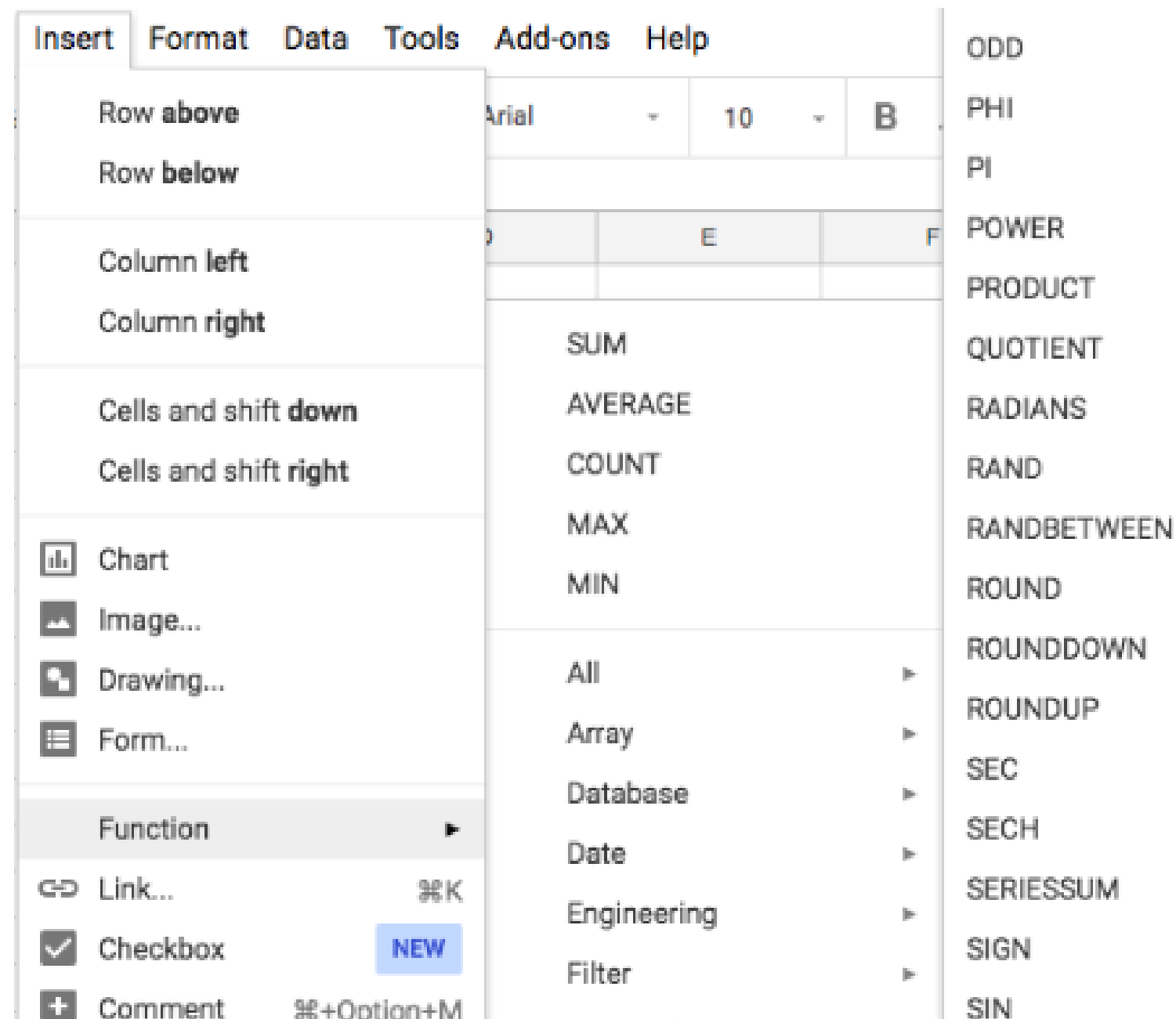


Chris Cardillo
Data Scientist at DataCamp

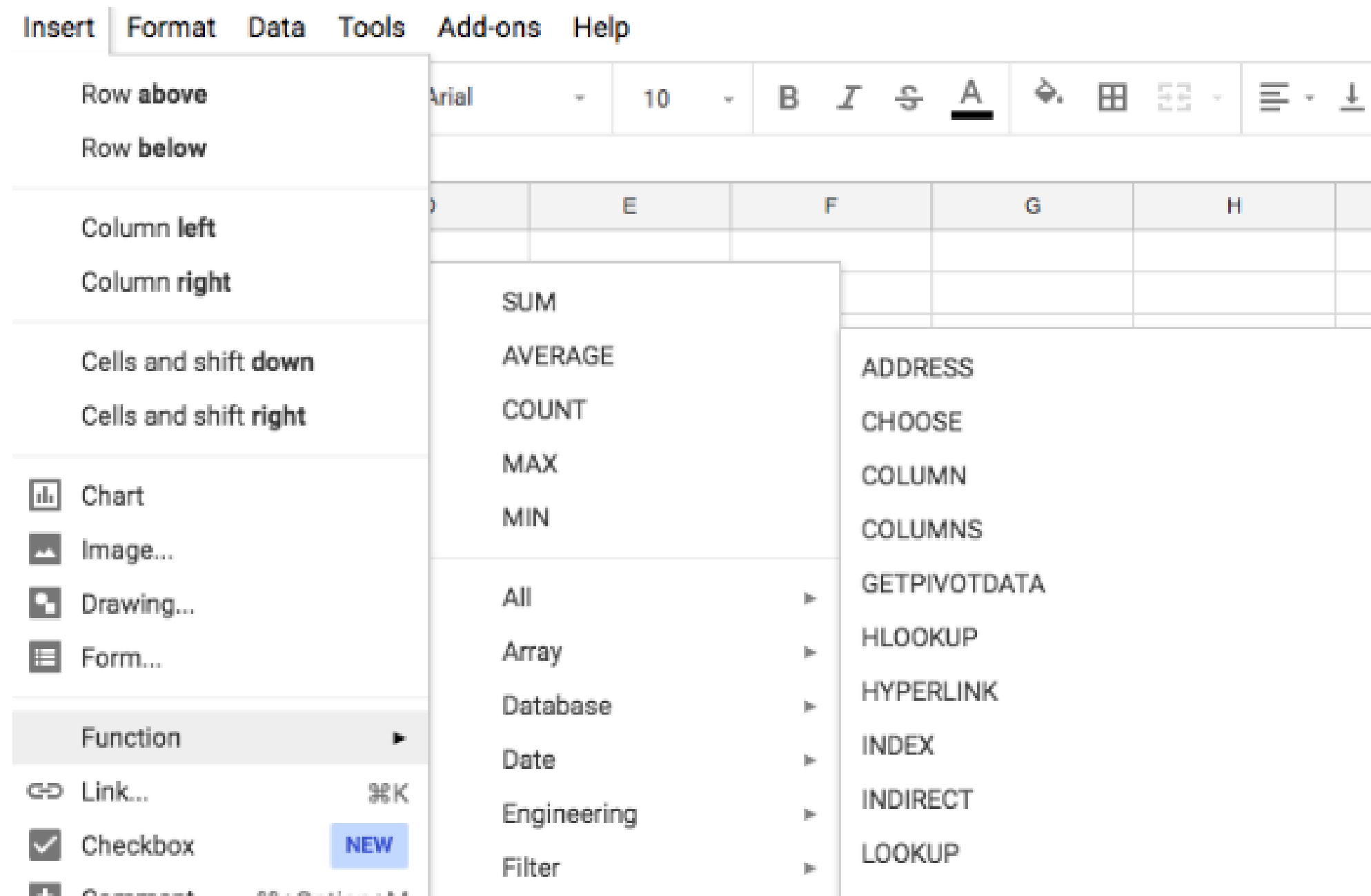
What is Python?

- General purpose programming language
- Has code packages for data science
- **Object-oriented**

Object-oriented spreadsheets



Object-oriented spreadsheets



Object-oriented Python

In Python

```
math.SUM(2, 3)
```

In Spreadsheets



Example Sheet



File Edit View Insert Format



100%



\$

%

.0
+

fx

=SUM(2, 3)

	A	B	C
1	5		

The power of the dot

```
math.SUM()
```

- `math` is our package
- `SUM` is our function in the package
- The dot (`.`) lets us access the function inside the package

Scripts

A Recipe for Buttered Toast

1. Turn on toaster
2. Put bread in toaster for 1 minute
3. Take bread out of toaster
4. Put butter on toast

Importing packages

```
import math
```


Importing pandas



Importing pandas

```
import pandas
```

Loading Excel data

```
import pandas

pandas.read_excel('fruit.xlsx')
```

	A	B	C
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

- fruit.xlsx
 - (the kind of data pandas likes)

pandas as pd

before

```
import pandas  
  
pandas.read_excel('fruit.xlsx')
```

after

```
import pandas as pd  
  
pd.read_excel('fruit.xlsx')
```

Variable assignment

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')
```

	A	B	C
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

fruit.xlsx

Looking at data with print()

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')

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

Script and console panes in DataCamp

SCRIPT.PY

```
1 # Import pandas library
2 import pandas as pd
3
4 # Import the data
5 fruit = pd.read_excel('fruit.xlsx')
6
7 # Print the data
8 print(fruit)|
```

- SCRIPT.PY is the script
- IPYTHON SHELL is the console



IPYTHON SHELL

SLIDES

```
name  color  price_usd
apple  red      0.50
```

Your turn!

PYTHON FOR SPREADSHEET USERS

DataFrames and their methods

PYTHON FOR SPREADSHEET USERS



Chris Cardillo
Data Scientist at DataCamp

Where we left off

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')

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

Anatomy of a pandas DataFrame

```
      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
- - - - -
```

Anatomy of a pandas DataFrame

	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	Blackberry	black	4.4

Anatomy of a pandas DataFrame

	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	Blackberry	black	1.42

Anatomy of a pandas DataFrame

	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

Anatomy of a pandas DataFrame

	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

Anatomy of a pandas DataFrame

	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

Anatomy of a pandas DataFrame

	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	Blackberry	black	1.14

DataFrame methods

- `.head()`
- `.info()`
- `.describe()`
- `.sort_values()`

The .head() method

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')

print(fruit.head())
```

	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

The .head() method

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')

print(fruit.head(2))
```

	name	color	price_usd
0	Apple	red	0.88
1	Banana	yellow	0.23

The .info() method

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')

print(fruit.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 3 columns):
name            8 non-null object
color           8 non-null object
price_usd       8 non-null float64
dtypes: float64(1), object(2)
memory usage: 272.0+ bytes
```

The .describe() method

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')

print(fruit.describe())
```

```
price_usd
count      8.000000
mean       2.285000
std        2.135911
min        0.230000
25%        0.830000
50%        1.000000
```

The .sort_values() method

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')

fruit = fruit.sort_values('name')
fruit = fruit.reset_index(drop=True)

print(fruit)
```

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

The .sort_values() method

```
import pandas as pd

fruit = pd.read_excel('fruit.xlsx')

fruit = fruit.sort_values('price_usd', ascending=False)
fruit = fruit.reset_index(drop=True)

print(fruit.head(3))
```

	name	color	price_usd
0	Dragonfruit	pink	5.27
1	Blueberries	blue	5.16

Your turn!

PYTHON FOR SPREADSHEET USERS

Filtering rows and creating columns

PYTHON FOR SPREADSHEET USERS



Chris Cardillo
Data Scientist at DataCamp

Filtering in spreadsheets

	A	B	C	D	
1	name ▼	color ▾	price_usd ▾		
2	Apple				
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					

Sort A → Z

Sort Z → A

▶ Filter by condition...

▼ Filter by values...

[Select all](#) - [Clear](#)

✓ Apple

Banana

Blueberries

Accessing a single column

```
      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
_ _ _ _ _
```

Accessing a single column

```
fruit['name']
```

```
0      Apple
1     Banana
2     Orange
3  Watermelon
4       Plum
5  Blueberries
6  Dragonfruit
7        Kiwi
Name: name, dtype: object
```

Comparison operators

- `==` (equal to)
- `!=` (not equal to)
- `<` (less than)
- `>` (greater than)
- `<=` (less than or equal to)
- `>=` (greater or equal to)

Comparisons

```
fruit['name']
```

0	Apple
1	Banana
2	Orange
3	Watermelon
4	Plum
5	Blueberries
6	Dragonfruit
7	Kiwi

```
fruit['name'] == "Apple"
```

0	True
1	False
2	False
3	False
4	False
5	False
6	False
7	False

Filtering

```
fruit[fruit['name'] == "Apple"]
```

	name	color	price_usd
0	Apple	red	0.88

Filtering

```
fruit[fruit['price_usd'] > 1]
```

	name	color	price_usd
3	Watermelon	green	3.98
5	Blueberries	blue	5.16
6	Dragonfruit	pink	5.27
7	Kiwi	brown	1.12

Filtering

```
fruit[fruit['price_usd'] > 1].reset_index(drop=True)
```

	name	color	price_usd
0	Watermelon	green	3.98
1	Blueberries	blue	5.16
2	Dragonfruit	pink	5.27
3	Kiwi	brown	1.12

Basic filtering pattern

```
df[df['this_column'] == 'that value']
```

Basic filtering pattern

In Python

```
fruit[fruit['name'] == "Apple"]
```

```
      name  color  price_usd
0  Apple    red      0.88
```

In Spreadsheets

	A	B	C	D	
1	name	color	price_usd		
2	Apple				
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					

Sort A → Z

Sort Z → A

Filter by condition...

Filter by values...

[Select all](#) - [Clear](#)

✓ Apple

Banana

Blueberries

Creating a new column

```
      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
_  _  _  _  _
```

Creating a new column

<i>fx</i>	=C5*2			
	A	B	C	D
1	name	color	price_usd	cost
2	Apple	red	0.88	1.76
3	Banana	yellow	0.23	0.46
4	Orange	orange	0.68	1.36
5	Watermelon	green	3.98	<div>7.96 ×</div> <div>=C5*2</div>
6	Plum	purple	0.96	1.92

Mathematical operators

- `+` (addition)
- `-` (subtraction)
- `*` (multiplication)
- `/` (division)

Creating a new column

```
fruit['cost'] = fruit['price_usd'] * 2
```

	name	color	price_usd	cost
0	Apple	red	0.88	1.76
1	Banana	yellow	0.23	0.46
2	Orange	orange	0.68	1.36
3	Watermelon	green	3.98	7.96
4	Plum	purple	0.96	1.92
-	-	-	-	-

Creating a new column

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

Creating a new column

fx | `=D5*C5`

	A	B	C	D	E
1	name	color	price_usd	quantity	cost
2	Apple	red	0.88	3	2.64
3	Banana	yellow	0.23	2	0.46
4	Orange	orange	0.68	2	1.36
5	Watermelon	green	3.98	1	<code>=D5*C5</code>
6	Plum	purple	0.96	0	0
7	Blueberries	blue	5.16	1	5.16

3.98 ×

`=D5*C5`

Creating a new column

```
fruit['cost'] = fruit['price_usd'] * fruit['quantity']
```

	name	color	price_usd	quantity	cost
0	Apple	red	0.88	3	2.64
1	Banana	yellow	0.23	2	0.46
2	Orange	orange	0.68	2	1.36
3	Watermelon	green	3.98	1	3.98
4	Plum	purple	0.96	0	0.00
5	Blueberries	blue	5.16	1	5.16
6	Dragonfruit	pink	5.27	1	5.27

Your turn!

PYTHON FOR SPREADSHEET USERS