Grouping and summing: the beginner's pivot table

PYTHON FOR SPREADSHEET USERS

Chris Cardillo
Data Scientist at DataCamp

Fruit stores

	A	B	C	D
1	store	product_name	quantity_purchased	revenue
2	Pete's Discount Fruit	Banana	1	0.23
3	Derek's Fruit Stand	Banana	3	0.69
4	Pete's Discount Fruit	Orange	1	0.68
5	Derek's Fruit Stand	Orange	2	1.36
6	Pete's Discount Fruit	Apple	1	0.88
7	Derek's Fruit Stand	Apple	1	0.88
8	Derek's Fruit Stand	Plum	1	0.96

Fruit stores

	A	B	C	D
1	store	product_name	quantity_purchased	revenue
2	Pete's Discount Fruit	Banana	1	0.23
3	Derek's Fruit Stand	Banana	3	0.69
4	Pete's Discount Fruit	Orange	1	0.68
5	Derek's Fruit Stand	Orange	2	1.36
6	Pete's Discount Fruit	Apple	1	0.88
7	Derek's Fruit Stand	Apple	1	0.88
8	Derek's Fruit Stand	Plum	1	0.96

Fruit stores

	A	B	C	D
1	store	product_name	quantity_purchased	revenue
2	Pete's Discount Fruit	Banana	1	0.23
3	Derek's Fruit Stand	Banana	3	0.69
4	Pete's Discount Fruit	Orange	1	0.68
5	Derek's Fruit Stand	Orange	2	1.36
6	Pete's Discount Fruit	Apple	1	0.88
7	Derek's Fruit Stand	Apple	1	0.88
8	Derek's Fruit Stand	Plum	1	0.96

Fruit stores

	A	B	C	D
1	store	product_name	quantity_purchased	revenue
2	Pete's Discount Fruit	Banana	1	0.23
3	Derek's Fruit Stand	Banana	3	0.69
4	Pete's Discount Fruit	Orange	1	0.68
5	Derek's Fruit Stand	Orange	2	1.36
6	Pete's Discount Fruit	Apple	1	0.88
7	Derek's Fruit Stand	Apple	1	0.88
8	Derek's Fruit Stand	Plum	1	0.96

Fruit stores

	A	B	C	D
1	store	product_name	quantity_purchased	revenue
2	Pete's Discount Fruit	Banana	1	0.23
3	Derek's Fruit Stand	Banana	3	0.69
4	Pete's Discount Fruit	Orange	1	0.68
5	Derek's Fruit Stand	Orange	2	1.36
6	Pete's Discount Fruit	Apple	1	0.88
7	Derek's Fruit Stand	Apple	1	0.88
8	Derek's Fruit Stand	Plum	1	0.96

Pivoting in spreadsheets

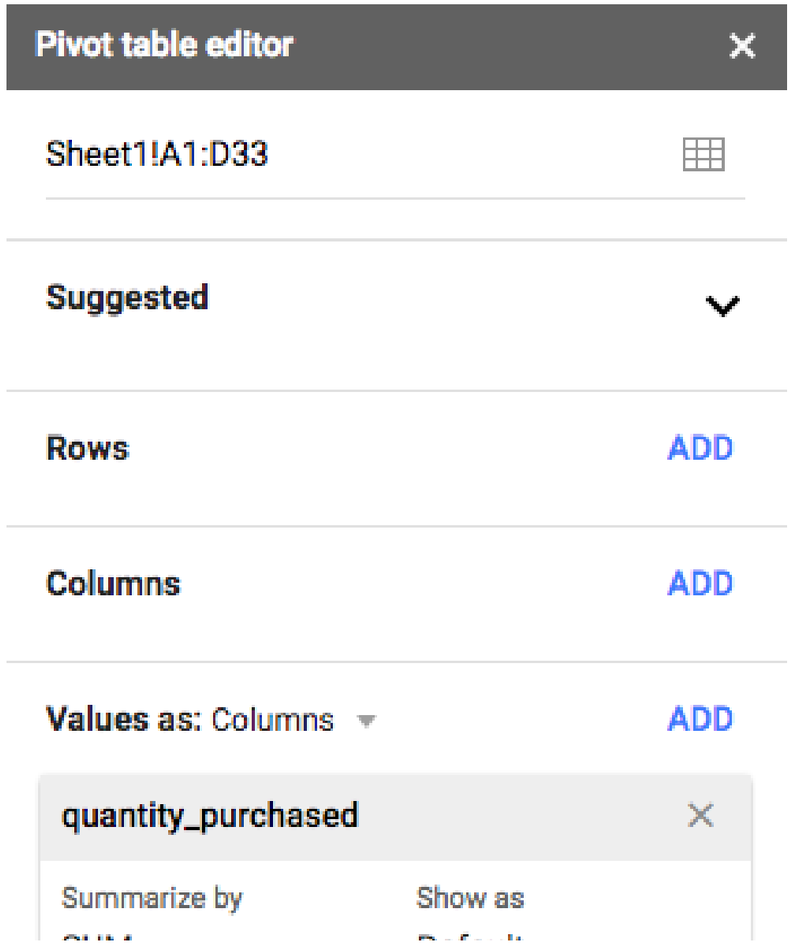
RAW DATA

	A	B	C	D
1	store	product_name	quantity_purchased	revenue
2	Pete's Discount Fruit	Banana	1	0.23
3	Derek's Fruit Stand	Banana	3	0.69
4	Pete's Discount Fruit	Orange	1	0.68
5	Derek's Fruit Stand	Orange	2	1.36
6	Pete's Discount Fruit	Apple	1	0.88
7	Derek's Fruit Stand	Apple	1	0.88
8	Derek's Fruit Stand	Plum	1	0.96
9	Derek's Fruit Stand	Plum	2	1.92
10	Derek's Fruit Stand	Kiwi	2	2.24
11	Derek's Fruit Stand	Watermelon	1	3.98

PIVOT TABLE

	A	B

PIVOT TABLE EDITOR



Pivoting in spreadsheets

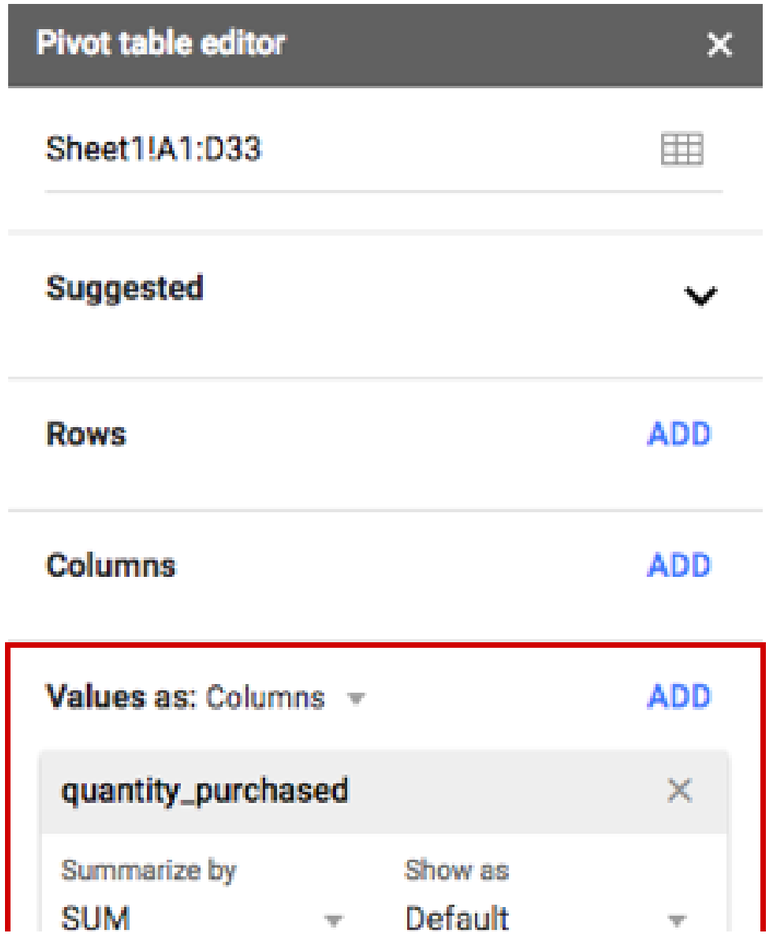
RAW DATA

	A	B	C	D
1	store	product_name	quantity_purchased	revenue
2	Pete's Discount Fruit	Banana	1	0.23
3	Derek's Fruit Stand	Banana	3	0.69
4	Pete's Discount Fruit	Orange	1	0.68
5	Derek's Fruit Stand	Orange	2	1.36
6	Pete's Discount Fruit	Apple	1	0.88
7	Derek's Fruit Stand	Apple	1	0.88
8	Derek's Fruit Stand	Plum	1	0.96
9	Derek's Fruit Stand	Plum	2	1.92
10	Derek's Fruit Stand	Kiwi	2	2.24
11	Derek's Fruit Stand	Watermelon	1	3.98

PIVOT TABLE

	A	B

PIVOT TABLE EDITOR



.sum()

```
fruit_sales.head()
```

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Banana	1	0.23
1	Derek's Fruit Stand	Banana	3	0.69
2	Pete's Discount Fruit	Orange	1	0.68
3	Derek's Fruit Stand	Orange	2	1.36
4	Pete's Discount Fruit	Apple	1	0.88

```
fruit_sales.sum(numeric_only=True)
```

```
quantity_purchased     57.00
revenue               133.12
dtype: float64
```

Pure summary with .sum()

IN SPREADSHEETS

	A	B
1	SUM of quantity_purchased	SUM of revenue
2	57	133.12

IN PYTHON

```
fruit_sales.sum(numeric_only=True)
```

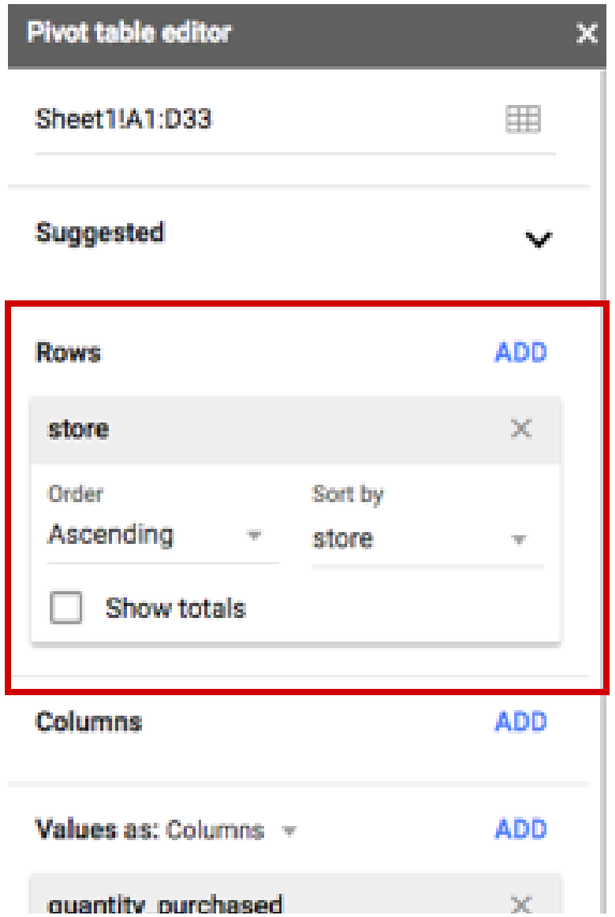
```
quantity_purchased     57.00
revenue                133.12
dtype: float64
```

Using pivot table rows in spreadsheets

PIVOT TABLE

	A	B	C
1	*store*	SUM of quantity_purchased	SUM of revenue
2	Derek's Fruit Stand	27	57.22
3	Pete's Discount Fruit	30	75.9

PIVOT TABLE EDITOR



A simple pivot table in Python - .groupby().sum()

In Python

```
fruit_sales.groupby('store', as_index=False).sum()
```

	store	quantity_purchased	revenue
0	Derek's Fruit Stand	27	57.22
1	Pete's Discount Fruit	30	75.90

In spreadsheets

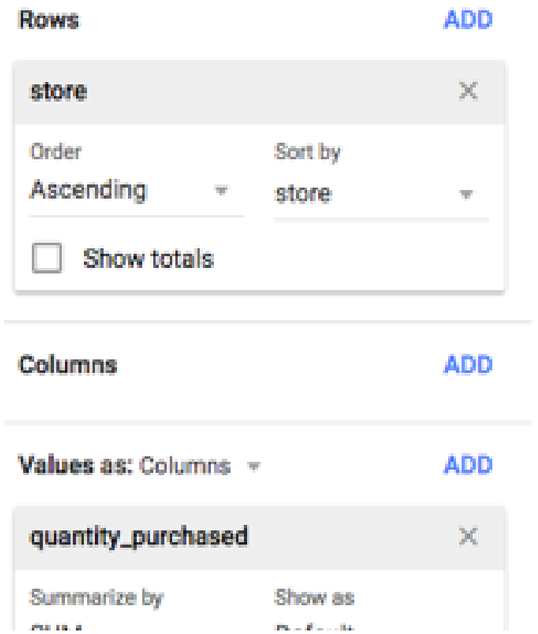
	A	B	C
1	store	SUM of quantity_purchased	SUM of revenue

A simple pivot table in Python - .groupby().sum()

In Python

```
fruit_sales.groupby('store', as_index=False).sum()
```

IN SPREADSHEETS PIVOT TABLE EDITOR



A simple pivot table in Python - .groupby().sum()

In Python

```
fruit_sales.groupby('store', as_index=False).sum()
```

	store	quantity_purchased	revenue
0	Derek's Fruit Stand	27	57.22
1	Pete's Discount Fruit	30	75.90

In spreadsheets

	A	B	C
1	store	SUM of quantity_purchased	SUM of revenue

Your turn!

PYTHON FOR SPREADSHEET USERS

Grouping by multiple columns

PYTHON FOR SPREADSHEET USERS

Chris Cardillo
Data Scientist at DataCamp

Fruit sales

`fruit_sales`

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Banana	1	0.23
1	Derek's Fruit Stand	Banana	3	0.69
2	Pete's Discount Fruit	Orange	1	0.68
3	Derek's Fruit Stand	Orange	2	1.36
4	Pete's Discount Fruit	Apple	1	0.88
5	Derek's Fruit Stand	Apple	1	0.88
6	Derek's Fruit Stand	Plum	1	0.96
7	Derek's Fruit Stand	Plum	2	1.92
8	Derek's Fruit Stand	Kiwi	2	2.24
9	Derek's Fruit Stand	Watermelon	1	3.98
10	Pete's Discount Fruit	Blueberries	3	15.48
11	Derek's Fruit Stand	Dragonfruit	3	15.81
12	Pete's Discount Fruit	Banana	1	0.23
13	Derek's Fruit Stand	Banana	1	0.23
14	Pete's Discount Fruit	Apple	3	2.64
15	Pete's Discount Fruit	Plum	2	1.92
16	Pete's Discount Fruit	Kiwi	3	3.36
17	Derek's Fruit Stand	Kiwi	3	3.36
18	Derek's Fruit Stand	Watermelon	2	7.96
19	Pete's Discount Fruit	Blueberries	3	15.48
20	Derek's Fruit Stand	Blueberries	1	5.16
21	Derek's Fruit Stand	Blueberries	1	5.16
22	Pete's Discount Fruit	Dragonfruit	3	15.81

`fruit_sales.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
Data columns (total 4 columns):
store                 32 non-null object
product_name          32 non-null object
quantity_purchased    32 non-null int64
revenue               32 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 1.1+ KB
```

Fruit sales

`fruit_sales`

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Banana	1	0.23
1	Derek's Fruit Stand	Banana	3	0.69
2	Pete's Discount Fruit	Orange	1	0.68
3	Derek's Fruit Stand	Orange	2	1.36
4	Pete's Discount Fruit	Apple	1	0.88
5	Derek's Fruit Stand	Apple	1	0.88
6	Derek's Fruit Stand	Plum	1	0.96
7	Derek's Fruit Stand	Plum	2	1.92
8	Derek's Fruit Stand	Kiwi	2	2.24
9	Derek's Fruit Stand	Watermelon	1	3.98
10	Pete's Discount Fruit	Blueberries	3	15.48
11	Derek's Fruit Stand	Dragonfruit	3	15.81
12	Pete's Discount Fruit	Banana	1	0.23
13	Derek's Fruit Stand	Banana	1	0.23
14	Pete's Discount Fruit	Apple	3	2.64
15	Pete's Discount Fruit	Plum	2	1.92
16	Pete's Discount Fruit	Kiwi	3	3.36
17	Derek's Fruit Stand	Kiwi	3	3.36
18	Derek's Fruit Stand	Watermelon	2	7.96
19	Pete's Discount Fruit	Blueberries	3	15.48
20	Derek's Fruit Stand	Blueberries	1	5.16
21	Derek's Fruit Stand	Blueberries	1	5.16
22	Pete's Discount Fruit	Dragonfruit	3	15.81

`fruit_sales.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
Data columns (total 4 columns):
store                 32 non-null object
product_name          32 non-null object
quantity_purchased    32 non-null int64
revenue               32 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 1.1+ KB
```

Fruit sales

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Banana	1	0.23
1	Derek's Fruit Stand	Banana	3	0.69
2	Pete's Discount Fruit	Orange	1	0.68
3	Derek's Fruit Stand	Orange	2	1.36
4	Pete's Discount Fruit	Apple	1	0.88
5	Derek's Fruit Stand	Apple	1	0.88
6	Derek's Fruit Stand	Plum	1	0.96
7	Derek's Fruit Stand	Plum	2	1.92
8	Derek's Fruit Stand	Kiwi	2	2.24
9	Derek's Fruit Stand	Watermelon	1	3.98
10	Pete's Discount Fruit	Blueberries	3	15.48
11	Derek's Fruit Stand	Dragonfruit	3	15.81
12	Pete's Discount Fruit	Banana	1	0.23
13	Derek's Fruit Stand	Banana	1	0.23
14	Pete's Discount Fruit	Apple	3	2.64
15	Pete's Discount Fruit	Plum	2	1.92

Adding a list of column names

BEFORE

```
fruit_sales.groupby('store', as_index=False).sum()
```

AFTER

```
fruit_sales.groupby(['store', 'product_name'], as_index=False).sum()
```

What is a list?

```
shopping_list = ['milk', 'eggs', 'cheese']
```

By store, by fruit

```
groups = ['store', 'product_name']

fruit_sales_less = fruit_sales.groupby(groups, as_index=False).sum()
```

	store	product_name	quantity_purchased	revenue
0	Derek's Fruit Stand	Apple	2	1.76
1	Derek's Fruit Stand	Banana	4	0.92
2	Derek's Fruit Stand	Blueberries	2	10.32
3	Derek's Fruit Stand	Dragonfruit	4	21.08
4	Derek's Fruit Stand	Kiwi	5	5.60
5	Derek's Fruit Stand	Orange	4	2.72
6	Derek's Fruit Stand	Plum	3	2.88
7	Derek's Fruit Stand	Watermelon	3	11.94
8	Pete's Discount Fruit	Apple	4	3.52
9	Pete's Discount Fruit	Banana	2	0.46
10	Pete's Discount Fruit	Blueberries	6	30.96

By store, by fruit

```
groups = ['store', 'product_name']

fruit_sales_less = fruit_sales.groupby(groups, as_index=False).sum()
```

	store	product_name	quantity_purchased	revenue
0	Derek's Fruit Stand	Apple	2	1.76
1	Derek's Fruit Stand	Banana	4	0.92
2	Derek's Fruit Stand	Blueberries	2	10.32
3	Derek's Fruit Stand	Dragonfruit	4	21.08
4	Derek's Fruit Stand	Kiwi	5	5.60
5	Derek's Fruit Stand	Orange	4	2.72
6	Derek's Fruit Stand	Plum	3	2.88
7	Derek's Fruit Stand	Watermelon	3	11.94
8	Pete's Discount Fruit	Apple	4	3.52
9	Pete's Discount Fruit	Banana	2	0.46
10	Pete's Discount Fruit	Blueberries	6	30.96
11	Pete's Discount Fruit	Dragonfruit	4	21.08

The benefits of grouping by more columns before .sum()

- It's not "one or none"
- Reduce data down to what matters
- Help make spreadsheet data more manageable

Your turn!

PYTHON FOR SPREADSHEET USERS

More Ways to Condense Information

PYTHON FOR SPREADSHEET USERS

Chris Cardillo
Data Scientist at DataCamp

.mean()

```
fruit_sales.groupby('store', as_index=False).mean()
```

	store	quantity_purchased	revenue
0	Derek's Fruit Stand	1.6875	3.57625
1	Pete's Discount Fruit	1.8750	4.74375

Steps to an answer

1. Fruit store transactions
2. Total sales by fruit for each store
3. Sorted sales by fruit for each store
4. Top row for each store

1: Fruit store transactions

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Banana	1	0.23
1	Derek's Fruit Stand	Banana	3	0.69
2	Pete's Discount Fruit	Orange	1	0.68
3	Derek's Fruit Stand	Orange	2	1.36
4	Pete's Discount Fruit	Apple	1	0.88
5	Derek's Fruit Stand	Apple	1	0.88
6	Derek's Fruit Stand	Plum	1	0.96
7	Derek's Fruit Stand	Plum	2	1.92
8	Derek's Fruit Stand	Kiwi	2	2.24
9	Derek's Fruit Stand	Watermelon	1	3.98
10	Pete's Discount Fruit	Blueberries	3	15.48
11	Derek's Fruit Stand	Dragonfruit	3	15.81
12	Pete's Discount Fruit	Banana	1	0.23
13	Derek's Fruit Stand	Banana	1	0.23
14	Pete's Discount Fruit	Apple	3	2.64
15	Pete's Discount Fruit	Plum	2	1.92

2: Total sales by fruit for each store

```
totals = fruit_sales.groupby(['store', 'product_name'], as_index=False).sum()
```

	store	product_name	quantity_purchased	revenue
0	Derek's Fruit Stand	Apple	2	1.76
1	Derek's Fruit Stand	Banana	4	0.92
2	Derek's Fruit Stand	Blueberries	2	10.32
3	Derek's Fruit Stand	Dragonfruit	4	21.08
4	Derek's Fruit Stand	Kiwi	5	5.60
5	Derek's Fruit Stand	Orange	4	2.72
6	Derek's Fruit Stand	Plum	3	2.88
7	Derek's Fruit Stand	Watermelon	3	11.94
8	Pete's Discount Fruit	Apple	4	3.52
9	Pete's Discount Fruit	Banana	2	0.46
10	Pete's Discount Fruit	Blueberries	6	30.96

3: Sorted sales by fruit for each store

```
totals = (totals.sort_values('revenue', ascending=False)
.reset_index(drop=True))
```

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Blueberries	6	30.96
1	Derek's Fruit Stand	Dragonfruit	4	21.08
2	Pete's Discount Fruit	Dragonfruit	4	21.08
3	Derek's Fruit Stand	Watermelon	3	11.94
4	Derek's Fruit Stand	Blueberries	2	10.32
5	Pete's Discount Fruit	Watermelon	2	7.96
6	Pete's Discount Fruit	Kiwi	6	6.72
7	Derek's Fruit Stand	Kiwi	5	5.60
8	Pete's Discount Fruit	Plum	4	3.84
9	Pete's Discount Fruit	Apple	4	3.52
10	Derek's Fruit Stand	Plum	3	2.88

4: Top row for each store

```
top_store_sellers = totals.groupby('store').head(1).reset_index(drop=True)
```

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Blueberries	6	30.96
1	Derek's Fruit Stand	Dragonfruit	4	21.08

Steps to an answer

```
# Raw data
fruit_sales

# Summary - by fruit by store
totals = fruit_sales.groupby(['store', 'product_name'], as_index=False).sum()

# Sort the Summary
totals = (totals.sort_values('revenue', ascending=False)
.reset_index(drop=True))

# First row for each store
top_store_sellers = totals.groupby('store').head(1).reset_index(drop=True)
```

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Blueberries	6	30.96
1	Derek's Fruit Stand	Dragonfruit	4	21.08

Steps to an answer

1: RAW TRANSACTION DATA

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Banana	1	0.23
1	Derek's Fruit Stand	Banana	3	0.69
2	Pete's Discount Fruit	Orange	1	0.68
3	Derek's Fruit Stand	Orange	2	1.36
4	Pete's Discount Fruit	Apple	1	0.88
5	Derek's Fruit Stand	Apple	1	0.88
6	Derek's Fruit Stand	Plum	1	0.96
7	Derek's Fruit Stand	Plum	2	1.92
8	Derek's Fruit Stand	Kiwi	2	2.24
9	Derek's Fruit Stand	Watermelon	1	3.98
10	Pete's Discount Fruit	Blueberries	3	15.48
11	Derek's Fruit Stand	Dragonfruit	3	15.81
12	Pete's Discount Fruit	Banana	1	0.23
13	Derek's Fruit Stand	Banana	1	0.23
14	Pete's Discount Fruit	Apple	3	2.64
15	Pete's Discount Fruit	Plum	2	1.92
16	Pete's Discount Fruit	Kiwi	3	3.36
17	Derek's Fruit Stand	Kiwi	3	3.36
18	Derek's Fruit Stand	Watermelon	2	7.96
19	Pete's Discount Fruit	Blueberries	3	15.48
20	Derek's Fruit Stand	Blueberries	1	5.16
21	Derek's Fruit Stand	Blueberries	1	5.16
22	Pete's Discount Fruit	Dragonfruit	3	15.81

2: BY FRUIT/STORE SUMMARY

	store	product_name	quantity_purchased	revenue
0	Derek's Fruit Stand	Apple	2	1.76
1	Derek's Fruit Stand	Banana	4	0.92
2	Derek's Fruit Stand	Blueberries	2	10.32
3	Derek's Fruit Stand	Dragonfruit	4	21.08
4	Derek's Fruit Stand	Kiwi	5	5.60
5	Derek's Fruit Stand	Orange	4	2.72
6	Derek's Fruit Stand	Plum	3	2.88
7	Derek's Fruit Stand	Watermelon	3	11.94
8	Pete's Discount Fruit	Apple	4	3.52
9	Pete's Discount Fruit	Banana	2	0.46
10	Pete's Discount Fruit	Blueberries	6	30.96
11	Pete's Discount Fruit	Dragonfruit	4	21.08
12	Pete's Discount Fruit	Kiwi	6	6.72
13	Pete's Discount Fruit	Orange	2	1.36
14	Pete's Discount Fruit	Plum	4	3.84
15	Pete's Discount Fruit	Watermelon	2	7.96

Steps to an answer

2: BY FRUIT/STORE SUMMARY

	store	product_name	quantity_purchased	revenue
0	Derek's Fruit Stand	Apple	2	1.76
1	Derek's Fruit Stand	Banana	4	0.92
2	Derek's Fruit Stand	Blueberries	2	10.32
3	Derek's Fruit Stand	Dragonfruit	4	21.08
4	Derek's Fruit Stand	Kiwi	5	5.60
5	Derek's Fruit Stand	Orange	4	2.72
6	Derek's Fruit Stand	Plum	3	2.88
7	Derek's Fruit Stand	Watermelon	3	11.94
8	Pete's Discount Fruit	Apple	4	3.52
9	Pete's Discount Fruit	Banana	2	0.46
10	Pete's Discount Fruit	Blueberries	6	30.96
11	Pete's Discount Fruit	Dragonfruit	4	21.08
12	Pete's Discount Fruit	Kiwi	6	6.72
13	Pete's Discount Fruit	Orange	2	1.36
14	Pete's Discount Fruit	Plum	4	3.84
15	Pete's Discount Fruit	Watermelon	2	7.96

3: BY FRUIT/STORE SORTED

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Blueberries	6	30.96
1	Derek's Fruit Stand	Dragonfruit	4	21.08
2	Pete's Discount Fruit	Dragonfruit	4	21.08
3	Derek's Fruit Stand	Watermelon	3	11.94
4	Derek's Fruit Stand	Blueberries	2	10.32
5	Pete's Discount Fruit	Watermelon	2	7.96
6	Pete's Discount Fruit	Kiwi	6	6.72
7	Derek's Fruit Stand	Kiwi	5	5.60
8	Pete's Discount Fruit	Plum	4	3.84
9	Pete's Discount Fruit	Apple	4	3.52
10	Derek's Fruit Stand	Plum	3	2.88
11	Derek's Fruit Stand	Orange	4	2.72
12	Derek's Fruit Stand	Apple	2	1.76
13	Pete's Discount Fruit	Orange	2	1.36
14	Derek's Fruit Stand	Banana	4	0.92
15	Pete's Discount Fruit	Banana	2	0.46

Steps to an answer

3: BY FRUIT/STORE SORTED

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Blueberries	6	30.96
1	Derek's Fruit Stand	Dragonfruit	4	21.08
2	Pete's Discount Fruit	Dragonfruit	4	21.08
3	Derek's Fruit Stand	Watermelon	3	11.94
4	Derek's Fruit Stand	Blueberries	2	10.32
5	Pete's Discount Fruit	Watermelon	2	7.96
6	Pete's Discount Fruit	Kiwi	6	6.72
7	Derek's Fruit Stand	Kiwi	5	5.60
8	Pete's Discount Fruit	Plum	4	3.84
9	Pete's Discount Fruit	Apple	4	3.52
10	Derek's Fruit Stand	Plum	3	2.88
11	Derek's Fruit Stand	Orange	4	2.72
12	Derek's Fruit Stand	Apple	2	1.76
13	Pete's Discount Fruit	Orange	2	1.36
14	Derek's Fruit Stand	Banana	4	0.92
15	Pete's Discount Fruit	Banana	2	0.46

4: FIRST ROW PER STORE

	store	product_name	quantity_purchased	revenue
0	Pete's Discount Fruit	Blueberries	6	30.96
1	Derek's Fruit Stand	Dragonfruit	4	21.08

Your turn!

PYTHON FOR SPREADSHEET USERS