Introduction

Selecting specific values of a pandas DataFrame or Series to work on is an implicit step in almost any data operation you'll run, so one of the first things you need to learn in working with data in Python is how to go about selecting the data points relevant to you quickly and effectively.

In [1]:
```
import pandas as pd
reviews = pd.read_csv("datasets/winemag-data-130k-v2.csv", index_col=0)
```

Native accessors

Native Python objects provide good ways of indexing data. Pandas carries all of these over, which helps make it easy to start with.

Consider this DataFrame:

In [2]:
```
reviews
```

Out[2]:

	country	description	designation	points	price	province	region_1	regior
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	N
1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	N
2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willame Va
3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	N
4	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	Willame Va
...	...	...	...	...	...	...	...	
65494	France	Made from young vines from the Vaulorent porti...	Fourchaume Premier Cru	90	45.0	Burgundy	Chablis	N
65495	Australia	This is a big, fat, almost sweet-tasting Caber...	NaN	90	22.0	South Australia	McLaren Vale	N
65496	US	Much improved over the unripe 2005, Fritz's 20...	Estate	90	20.0	California	Dry Creek Valley	Sonc
65497	US	This wine wears its 15.8% alcohol	Block 24	90	31.0	California	Napa Valley	Na

	country	description	designation	points	price	province	region_1	regior
		better than ...						
65498	Spain	A unique take on Manzanilla Sherry, which is o...	Manzanilla	90	10.0	Andalucia	Jerez	N

65499 rows × 13 columns

In Python, we can access the property of an object by accessing it as an attribute. A `book` object, for example, might have a `title` property, which we can access by calling `book.title`. Columns in a pandas DataFrame work in much the same way.

Hence to access the `country` property of `reviews` we can use:

การเลือกคอลัมน์จาก DataFrame

- แสดงค่าในคอลัมน์เป็น Series

In [3]:
```
reviews.country
```

Out[3]:
```
0           Italy
1        Portugal
2              US
3              US
4              US
           ...
65494       France
65495    Australia
65496           US
65497           US
65498        Spain
Name: country, Length: 65499, dtype: object
```

If we have a Python dictionary, we can access its values using the indexing (`[]`) operator. We can do the same with columns in a DataFrame:

In [4]:
```
reviews['country']
```

Out[4]:
```
0            Italy
1         Portugal
2               US
3               US
4               US
            ...
65494       France
65495    Australia
65496           US
65497           US
65498        Spain
Name: country, Length: 65499, dtype: object
```

- แสดงค่าในคอลัมน์เป็น DataFrame

In [5]:
```
reviews[['country']]
```

Out[5]:

	country
0	Italy
1	Portugal
2	US
3	US
4	US
...	...
65494	France
65495	Australia
65496	US
65497	US
65498	Spain

65499 rows × 1 columns

In [6]:
```
my_cols = ['country','points','province','taster_name']
reviews[my_cols]
```

Out[6]:

	country	points	province	taster_name
0	Italy	87	Sicily & Sardinia	Kerin O'Keefe
1	Portugal	87	Douro	Roger Voss
2	US	87	Oregon	Paul Gregutt
3	US	87	Michigan	Alexander Peartree
4	US	87	Oregon	Paul Gregutt
...	...	...	...	...
65494	France	90	Burgundy	Roger Voss
65495	Australia	90	South Australia	Joe Czerwinski
65496	US	90	California	NaN
65497	US	90	California	NaN
65498	Spain	90	Andalucia	Michael Schachner

65499 rows × 4 columns

These are the two ways of selecting a specific Series out of a DataFrame. Neither of them is more or less syntactically valid than the other, but the indexing operator `[]` does have the advantage that it can handle column names with reserved characters in them (e.g. if we had a `country providence` column, `reviews.country providence` wouldn't work).

Doesn't a pandas Series look kind of like a fancy dictionary? It pretty much is, so it's no surprise that, to drill down to a single specific value, we need only use the indexing operator `[]` once more:

การเลือกตำแหน่งจาก DataFrame

In [7]:
```
reviews.country[0]
```

Out[7]:
```
'Italy'
```

In [8]:
```
reviews['country'][0]
```

Out[8]:
```
'Italy'
```

Indexing in pandas

The indexing operator and attribute selection are nice because they work just like they do in the rest of the Python ecosystem. As a novice, this makes them easy to pick up

and use. However, pandas has its own accessor operators, `loc` and `iloc`. For more advanced operations, these are the ones you're supposed to be using.

การเลือกแถว และ คอลัมน์

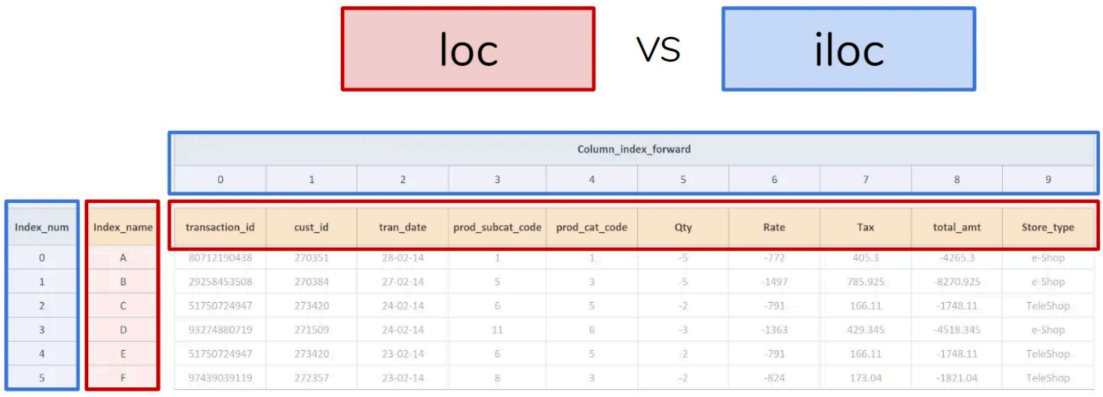
การที่จะเลือกแถว และคอลัมน์ นั้นเราจะต้องระบุว่าต้องการจะเลือกข้อมูลที่อยู่ใน แถว อะไร และคอลัมน์อะไร หรือ Location ที่มีข้อมูลที่ต้องการอยู่นั้นเอง ซึ่งมี 2 วิธี ดังนี้

1. loc : การที่จะใช้ loc นั้น จะต้องระบุชื่อของแถวและ ชื่อของคอลัมน์ลงไปตรงๆ

```
# dataframe.loc['index_name' , 'column_name']
```

2. iloc : การที่จะใช้ iloc นั้นไม่สนใจชื่อ แต่สนใจตำแหน่งของแถว และ ตำแหน่งของคอลัมน์

```
dataframe.iloc['row_index_number' , 'column_index_number']
```



Index-based selection

Pandas indexing works in one of two paradigms. The first is **index-based selection**: selecting data based on its numerical position in the data. `iloc` follows this paradigm.

To select the first row of data in a DataFrame, we may use the following:

```
In [9]: reviews.iloc[0]  #ถ้าแถวเดียวจะแสดงผลเป็น Series
```

Out[9]:
```
country                                                    Italy
description             Aromas include tropical fruit, broom, brimston...
designation                                          Vulkà Bianco
points                                                         87
price                                                         NaN
province                                          Sicily & Sardinia
region_1                                                    Etna
region_2                                                     NaN
taster_name                                         Kerin O’Keefe
taster_twitter_handle                                  @kerinokeefe
title                                 Nicosia 2013 Vulkà Bianco  (Etna)
variety                                              White Blend
winery                                                   Nicosia
Name: 0, dtype: object
```

In [10]:
```
reviews.iloc[[0]] #ถ้าอยากให้แถวเดียวแสดงผลเป็น DataFrame ต้องใส่ [] คลุมอีกที
```

Out[10]:

	country	description	designation	points	price	province	region_1	region_2	tast
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	NaN	

In [11]:
```
reviews.iloc[0:2] #ถ้าหลายแถวจะแสดงผลเป็น DataFrame
```

Out[11]:

	country	description	designation	points	price	province	region_1	region_2	tast
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	NaN	
1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	NaN	R

Both `loc` and `iloc` are row-first, column-second. This is the opposite of what we do in native Python, which is column-first, row-second.

This means that it's marginally easier to retrieve rows, and marginally harder to get retrieve columns. To get a column with `iloc`, we can do the following:

In [12]:
```
reviews.iloc[:, 0]
```

Out[12]:
```
0           Italy
1        Portugal
2              US
3              US
4              US
           ...
65494       France
65495    Australia
65496           US
65497           US
65498        Spain
Name: country, Length: 65499, dtype: object
```

On its own, the `:` operator, which also comes from native Python, means "everything". When combined with other selectors, however, it can be used to indicate a range of values. For example, to select the `country` column from just the first, second, and third row, we would do:

In [13]:
```
reviews.iloc[:3, 0]
```

Out[13]:
```
0       Italy
1    Portugal
2          US
Name: country, dtype: object
```

Or, to select just the second and third entries, we would do:

In [14]:
```
reviews.iloc[1:3, 0]
```

Out[14]:
```
1    Portugal
2          US
Name: country, dtype: object
```

In [15]:
```
reviews.iloc[2:4, :5]
```

Out[15]:

	country	description	designation	points	price
2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0
3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0

It's also possible to pass a list:

In [16]:
```
reviews.iloc[[2, 3], [0,1,2,3,4]]
```

Out[16]:

	country	description	designation	points	price
2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0
3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0

Finally, it's worth knowing that negative numbers can be used in selection. This will start counting forwards from the *end* of the values. So for example here are the last five elements of the dataset.

In [17]:
```
reviews.iloc[-5:]
```

Out[17]:

	country	description	designation	points	price	province	region_1	region_2
65494	France	Made from young vines from the Vaulorent porti...	Fourchaume Premier Cru	90	45.0	Burgundy	Chablis	NaN
65495	Australia	This is a big, fat, almost sweet-tasting Caber...	NaN	90	22.0	South Australia	McLaren Vale	NaN
65496	US	Much improved over the unripe 2005, Fritz's 20...	Estate	90	20.0	California	Dry Creek Valley	Sonoma
65497	US	This wine wears its 15.8% alcohol better than ...	Block 24	90	31.0	California	Napa Valley	Napa
65498	Spain	A unique take on Manzanilla Sherry, which is o...	Manzanilla	90	10.0	Andalucia	Jerez	NaN

Label-based selection

The second paradigm for attribute selection is the one followed by the `loc` operator: **label-based selection**. In this paradigm, it's the data index value, not its position, which

For example, to get the first entry in `reviews`, we would now do the following:

In [18]:
```
reviews.loc[0, 'country']
```

Out[18]:
```
'Italy'
```

`iloc` is conceptually simpler than `loc` because it ignores the dataset's indices. When we use `iloc` we treat the dataset like a big matrix (a list of lists), one that we have to index into by position. `loc`, by contrast, uses the information in the indices to do its work. Since your dataset usually has meaningful indices, it's usually easier to do things using `loc` instead. For example, here's one operation that's much easier using `loc`:

In [19]:
```
reviews.loc[:, ['taster_name', 'taster_twitter_handle', 'points']]
```

Out[19]:

	taster_name	taster_twitter_handle	points
0	Kerin O'Keefe	@kerinokeefe	87
1	Roger Voss	@vossroger	87
2	Paul Gregutt	@paulgwine	87
3	Alexander Peartree	NaN	87
4	Paul Gregutt	@paulgwine	87
...	...	...	...
65494	Roger Voss	@vossroger	90
65495	Joe Czerwinski	@JoeCz	90
65496	NaN	NaN	90
65497	NaN	NaN	90
65498	Michael Schachner	@wineschach	90

65499 rows × 3 columns

Choosing between `loc` and `iloc`

When choosing or transitioning between `loc` and `iloc`, there is one "gotcha" worth keeping in mind, which is that the two methods use slightly different indexing schemes.

`iloc` uses the Python stdlib indexing scheme, where the first element of the range is included and the last one excluded. So `0:10` will select entries `0,...,9`. `loc`, meanwhile, indexes inclusively. So `0:10` will select entries `0,...,10`.

Why the change? Remember that loc can index any stdlib type: strings, for example. If we have a DataFrame with index values `Apples, ..., Potatoes, ...`, and we want to select "all the alphabetical fruit choices between Apples and Potatoes", then it's

a lot more convenient to index `df.loc['Apples':'Potatoes']` than it is to index something like `df.loc['Apples', 'Potatoet']` (`t` coming after `s` in the alphabet).

This is particularly confusing when the DataFrame index is a simple numerical list, e.g. `0,...,1000`. In this case `df.iloc[0:1000]` will return 1000 entries, while `df.loc[0:1000]` return 1001 of them! To get 1000 elements using `loc`, you will need to go one lower and ask for `df.loc[0:999]`.

Otherwise, the semantics of using `loc` are the same as those for `iloc`.

Manipulating the index

Label-based selection derives its power from the labels in the index. Critically, the index we use is not immutable. We can manipulate the index in any way we see fit.

The `set_index()` method can be used to do the job. Here is what happens when we `set_index` to the `title` field:

In [20]:
```
reviews.set_index("title")
```

Out[20]:

title	country	description	designation	points	price	province	region_1	
Nicosia 2013 Vulkà Bianco (Etna)	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	
Quinta dos Avidagos 2011 Avidagos Red (Douro)	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	
Rainstorm 2013 Pinot Gris (Willamette Valley)	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	W
St. Julian 2013 Reserve Late Harvest Riesling (Lake Michigan Shore)	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	
Sweet Cheeks 2012 Vintner's Reserve Wild Child Block Pinot Noir (Willamette Valley)	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	W
...	...	...	...	...	...	...	...	
William Fèvre 2005 Fourchaume Premier Cru (Chablis)	France	Made from young vines from the Vaulorent porti...	Fourchaume Premier Cru	90	45.0	Burgundy	Chablis	
Tapestry 2005 Cabernet Sauvignon (McLaren Vale)	Australia	This is a big, fat, almost sweet-tasting Caber...	NaN	90	22.0	South Australia	McLaren Vale	
Estate	US	Much improved	Estate	90	20.0	California	Dry Creek Valley	

title	country	description	designation	points	price	province	region_1
Sauvignon Blanc (Dry Creek Valley)		over the unripe 2005, Fritz's 20...					
Hendry 2004 Block 24 Primitivo (Napa Valley)	US	This wine wears its 15.8% alcohol better than ...	Block 24	90	31.0	California	Napa Valley
Bodegas Dios Baco S.L. NV Manzanilla Sherry (Jerez)	Spain	A unique take on Manzanilla Sherry, which is o...	Manzanilla	90	10.0	Andalucia	Jerez

65499 rows × 12 columns

This is useful if you can come up with an index for the dataset which is better than the current one.

Conditional selection

So far we've been indexing various strides of data, using structural properties of the DataFrame itself. To do *interesting* things with the data, however, we often need to ask questions based on conditions.

For example, suppose that we're interested specifically in better-than-average wines produced in Italy.

We can start by checking if each wine is Italian or not:

In [21]:
```
reviews.country == 'Italy'
```

Out[21]:
```
0         True
1        False
2        False
3        False
4        False
         ...
65494    False
65495    False
65496    False
65497    False
         False
Name: country, Length: 65499, dtype: bool
```

This operation produced a Series of `True` / `False` booleans based on the `country` of each record. This result can then be used inside of `loc` to select the relevant data:

In [22]:
```
reviews.loc[reviews.country == 'Italy']
```

Out[22]:

	country	description	designation	points	price	province	region_1	reg
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	
6	Italy	Here's a bright, informal red that opens with ...	Belsito	87	16.0	Sicily & Sardinia	Vittoria	
13	Italy	This is dominated by oak and oak-driven aromas...	Rosso	87	NaN	Sicily & Sardinia	Etna	
22	Italy	Delicate aromas recall white flower and citrus...	Ficiligno	87	19.0	Sicily & Sardinia	Sicilia	
24	Italy	Aromas of prune, blackcurrant, toast and oak c...	Aynat	87	35.0	Sicily & Sardinia	Sicilia	
...	...	...	...	...	...	...	...	
65466	Italy	Earthy truffle, porcini mushroom, herb and gam...	NaN	88	70.0	Tuscany	Brunello di Montalcino	
65474	Italy	Made of 70% Syrah, 15% Sangiovese and 15% Merl...	Taneto	88	25.0	Tuscany	Toscana	
65476	Italy	Rose, violet, sour berry and tilled earth arom...	Prugnolo	88	25.0	Tuscany	Rosso di Montepulciano	
65477	Italy	Made of 65% Merlot, 25% Cabernet Sauvignon, 5%...	Ruit Hora	88	30.0	Tuscany	Bolgheri	

	country	description	designation	points	price	province	region_1	reg
65478	Italy	Aromas suggesting French oak, coconut and spic...	NaN	88	36.0	Tuscany	Vino Nobile di Montepulciano	

10005 rows × 13 columns

In [23]:
```
reviews[reviews.country == 'Italy']
```

Out[23]:

	country	description	designation	points	price	province	region_1	reg
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	
6	Italy	Here's a bright, informal red that opens with ...	Belsito	87	16.0	Sicily & Sardinia	Vittoria	
13	Italy	This is dominated by oak and oak-driven aromas...	Rosso	87	NaN	Sicily & Sardinia	Etna	
22	Italy	Delicate aromas recall white flower and citrus...	Ficiligno	87	19.0	Sicily & Sardinia	Sicilia	
24	Italy	Aromas of prune, blackcurrant, toast and oak c...	Aynat	87	35.0	Sicily & Sardinia	Sicilia	
...	...	...	...	...	...	...	...	
65466	Italy	Earthy truffle, porcini mushroom, herb and gam...	NaN	88	70.0	Tuscany	Brunello di Montalcino	
65474	Italy	Made of 70% Syrah, 15% Sangiovese and 15% Merl...	Taneto	88	25.0	Tuscany	Toscana	
65476	Italy	Rose, violet, sour berry and tilled earth arom...	Prugnolo	88	25.0	Tuscany	Rosso di Montepulciano	
65477	Italy	Made of 65% Merlot, 25% Cabernet Sauvignon, 5%...	Ruit Hora	88	30.0	Tuscany	Bolgheri	

	country	description	designation	points	price	province	region_1	reg
65478	Italy	Aromas suggesting French oak, coconut and spic...	NaN	88	36.0	Tuscany	Vino Nobile di Montepulciano	

10005 rows × 13 columns

This DataFrame has ~20,000 rows. The original had ~130,000. That means that around 15% of wines originate from Italy.

We also wanted to know which ones are better than average. Wines are reviewed on a 80-to-100 point scale, so this could mean wines that accrued at least 90 points.

We can use the ampersand (`&`) to bring the two questions together:

In [24]:
```
reviews.loc[(reviews.country == 'Italy') & (reviews.points >= 90)]
```

Out[24]:

	country	description	designation	points	price	province	region_1	re
120	Italy	Slightly backward, particularly given the vint...	Bricco Rocche Prapó	92	70.0	Piedmont	Barolo	
130	Italy	At the first it was quite muted and subdued, b...	Bricco Rocche Brunate	91	70.0	Piedmont	Barolo	
133	Italy	Einaudi's wines have been improving lately, an...	NaN	91	68.0	Piedmont	Barolo	
135	Italy	The color is just beginning to show signs of b...	Sorano	91	60.0	Piedmont	Barolo	
140	Italy	A big, fat, luscious wine with plenty of toast...	Costa Bruna	90	26.0	Piedmont	Barbera d'Alba	
...	...	...	...	...	...	...	...	
65225	Italy	You'll love the dark intensity and generous ar...	Vigna Manapetra Riserva	93	58.0	Tuscany	Brunello di Montalcino	
65226	Italy	Brunello Madonna Nera is a new product (this w...	Madonna Nera	92	NaN	Tuscany	Brunello di Montalcino	
65362	Italy	This stunning single-vineyard selection is one...	La Rocca	95	31.0	Veneto	Soave Classico	
65365	Italy	Stunning and sophisticated, it leads with inte...	Sanct Valentin	94	40.0	Northeastern Italy	Alto Adige	

	country	description	designation	points	price	province	region_1	re
65399	Italy	Aesthetics and elegance are important values t...	Nectar Dei	91	65.0	Tuscany	Maremma	

3479 rows × 13 columns

In [25]:
```
reviews[(reviews.country == 'Italy') & (reviews.points >= 90)]
```

Out[25]:

	country	description	designation	points	price	province	region_1	re
120	Italy	Slightly backward, particularly given the vint...	Bricco Rocche Prapó	92	70.0	Piedmont	Barolo	
130	Italy	At the first it was quite muted and subdued, b...	Bricco Rocche Brunate	91	70.0	Piedmont	Barolo	
133	Italy	Einaudi's wines have been improving lately, an...	NaN	91	68.0	Piedmont	Barolo	
135	Italy	The color is just beginning to show signs of b...	Sorano	91	60.0	Piedmont	Barolo	
140	Italy	A big, fat, luscious wine with plenty of toast...	Costa Bruna	90	26.0	Piedmont	Barbera d'Alba	
...	...	...	...	...	...	...	...	
65225	Italy	You'll love the dark intensity and generous ar...	Vigna Manapetra Riserva	93	58.0	Tuscany	Brunello di Montalcino	
65226	Italy	Brunello Madonna Nera is a new product (this w...	Madonna Nera	92	NaN	Tuscany	Brunello di Montalcino	
65362	Italy	This stunning single-vineyard selection is one...	La Rocca	95	31.0	Veneto	Soave Classico	
65365	Italy	Stunning and sophisticated, it leads with inte...	Sanct Valentin	94	40.0	Northeastern Italy	Alto Adige	

	country	description	designation	points	price	province	region_1	re
65399	Italy	Aesthetics and elegance are important values t...	Nectar Dei	91	65.0	Tuscany	Maremma	

3479 rows × 13 columns

Suppose we'll buy any wine that's made in Italy *or* which is rated above average. For this we use a pipe (`|`):

In [26]:
```
reviews[(reviews.country == 'Italy') | (reviews.points >= 90)]
```

Out[26]:

	country	description	designation	points	price	province	region_1	region_
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	Na
6	Italy	Here's a bright, informal red that opens with ...	Belsito	87	16.0	Sicily & Sardinia	Vittoria	Na
13	Italy	This is dominated by oak and oak-driven aromas...	Rosso	87	NaN	Sicily & Sardinia	Etna	Na
22	Italy	Delicate aromas recall white flower and citrus...	Ficiligno	87	19.0	Sicily & Sardinia	Sicilia	Na
24	Italy	Aromas of prune, blackcurrant, toast and oak c...	Aynat	87	35.0	Sicily & Sardinia	Sicilia	Na
...	...	...	...	...	...	...	...	
65494	France	Made from young vines from the Vaulorent porti...	Fourchaume Premier Cru	90	45.0	Burgundy	Chablis	Na
65495	Australia	This is a big, fat, almost sweet-tasting Caber...	NaN	90	22.0	South Australia	McLaren Vale	Na
65496	US	Much improved over the unripe 2005, Fritz's 20...	Estate	90	20.0	California	Dry Creek Valley	Sonom
65497	US	This wine wears its 15.8% alcohol better than ...	Block 24	90	31.0	California	Napa Valley	Nap

	country	description	designation	points	price	province	region_1	region_
65498	Spain	A unique take on Manzanilla Sherry, which is o...	Manzanilla	90	10.0	Andalucia	Jerez	Na

31430 rows × 13 columns

Pandas comes with a few built-in conditional selectors, two of which we will highlight here.

The first is `isin`. `isin` is lets you select data whose value "is in" a list of values. For example, here's how we can use it to select wines only from Italy or France:

In [27]:
```
reviews.loc[reviews.country.isin(['Italy', 'France'])]
```

Out[27]:

	country	description	designation	points	price	province	region_1	region_2
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	NaN
6	Italy	Here's a bright, informal red that opens with ...	Belsito	87	16.0	Sicily & Sardinia	Vittoria	NaN
7	France	This dry and restrained wine offers spice in p...	NaN	87	24.0	Alsace	Alsace	NaN
9	France	This has great depth of flavor with its fresh ...	Les Natures	87	27.0	Alsace	Alsace	NaN
11	France	This is a dry wine, very spicy, with a tight, ...	NaN	87	30.0	Alsace	Alsace	NaN
...	...	...	...	...	...	...	...	...
65485	France	There's a fine balance here between minerality...	Montmains Premier Cru	90	40.0	Burgundy	Chablis	NaN
65486	France	Closed up and firm with a hint of vanilla, hon...	Domaine Long-Depaquit Les Bougnons Premier Cru	90	NaN	Burgundy	Chablis	NaN
65491	France	A big, toasty wine, full of ripe, delicious fr...	Fourchaume Vieilles Vignes Premier Cru	90	36.0	Burgundy	Chablis	NaN
65492	France	A rounded, fruity wine, packed with yellow pea...	Mont-de-Milieu Premier Cru	90	30.0	Burgundy	Chablis	NaN
[illegible]	[illegible]ance	Made from young vines	Fourchaume Premier Cru	90	45.0	Burgundy	Chablis	NaN

	country	description	designation	points	price	province	region_1	region_2
		from the Vaulorent porti...						

21179 rows × 13 columns

The second is `isnull` (and its companion `notnull`). These methods let you highlight values which are (or are not) empty (`NaN`). For example, to filter out wines lacking a price tag in the dataset, here's what we would do:

In [28]:
```
reviews.loc[reviews.price.isnull()]
```

Out[28]:

	country	description	designation	points	price	province	region_1	r
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	
13	Italy	This is dominated by oak and oak-driven aromas...	Rosso	87	NaN	Sicily & Sardinia	Etna	
30	France	Red cherry fruit comes laced with light tannin...	Nouveau	86	NaN	Beaujolais	Beaujolais-Villages	
31	Italy	Merlot and Nero d'Avola form the base for this...	Calanìca Nero d'Avola-Merlot	86	NaN	Sicily & Sardinia	Sicilia	
32	Italy	Part of the extended Calanìca series, this Gri...	Calanìca Grillo-Viognier	86	NaN	Sicily & Sardinia	Sicilia	
...	...	...	...	...	...	...	...	
65252	US	There's a bit of a green bean streak running t...	NaN	87	NaN	Washington	Washington	Wa
65281	Portugal	This wine is all about black and red berry fru...	Casa Américo	86	NaN	Dão	NaN	
65284	Chile	Minerally raw aromas brush up against harsh. T...	Tectonia	86	NaN	Leyda Valley	NaN	
65405	Austria	Intense, crisp acidity shines on this young wi...	Steinagrund Lagenreserve	91	NaN	Wagram-Donauland	NaN	

	country	description	designation	points	price	province	region_1	r
65486	France	Closed up and firm with a hint of vanilla, hon...	Domaine Long-Depaquit Les Bougnons Premier Cru	90	NaN	Burgundy	Chablis	

4670 rows × 13 columns

In [29]:
```
reviews.loc[reviews.price.notnull()]
```

Out[29]:

	country	description	designation	points	price	province	region_1	regior
1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	N
2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willame Va
3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	N
4	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	Willame Va
5	Spain	Blackberry and raspberry aromas show a typical...	Ars In Vitro	87	15.0	Northern Spain	Navarra	N
...	...	...	...	...	...	...	...	
65494	France	Made from young vines from the Vaulorent porti...	Fourchaume Premier Cru	90	45.0	Burgundy	Chablis	N
65495	Australia	This is a big, fat, almost sweet-tasting Caber...	NaN	90	22.0	South Australia	McLaren Vale	N
65496	US	Much improved over the unripe 2005, Fritz's 20...	Estate	90	20.0	California	Dry Creek Valley	Sonc
65497	US	This wine wears its 15.8% alcohol	Block 24	90	31.0	California	Napa Valley	Na

	country	description	designation	points	price	province	region_1	regior
		better than ...						
65498	Spain	A unique take on Manzanilla Sherry, which is o...	Manzanilla	90	10.0	Andalucia	Jerez	N

60829 rows × 13 columns

Assigning data

Going the other way, assigning data to a DataFrame is easy. You can assign either a constant value:

In [30]:
```
reviews['critic'] = 'everyone'
reviews
```

Out[30]:

	country	description	designation	points	price	province	region_1	regior
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	N
1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	N
2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willame Va
3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	N
4	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	Willame Va
...	...	...	...	...	...	...	...	
65494	France	Made from young vines from the Vaulorent porti...	Fourchaume Premier Cru	90	45.0	Burgundy	Chablis	N
65495	Australia	This is a big, fat, almost sweet-tasting Caber...	NaN	90	22.0	South Australia	McLaren Vale	N
65496	US	Much improved over the unripe 2005, Fritz's 20...	Estate	90	20.0	California	Dry Creek Valley	Sonc
65497	US	This wine wears its 15.8% alcohol	Block 24	90	31.0	California	Napa Valley	Na

	country	description	designation	points	price	province	region_1	regior
		better than ...						
65498	Spain	A unique take on Manzanilla Sherry, which is o...	Manzanilla	90	10.0	Andalucia	Jerez	N

65499 rows × 14 columns

Or with an iterable of values:

In [31]:
```
reviews['index_backwards'] = range(len(reviews), 0, -1)
reviews
```

Out[31]:

	country	description	designation	points	price	province	region_1	regior
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	N
1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	N
2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willame Va
3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	N
4	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	Willame Va
...	...	...	...	...	...	...	...	
65494	France	Made from young vines from the Vaulorent porti...	Fourchaume Premier Cru	90	45.0	Burgundy	Chablis	N
65495	Australia	This is a big, fat, almost sweet-tasting Caber...	NaN	90	22.0	South Australia	McLaren Vale	N
65496	US	Much improved over the unripe 2005, Fritz's 20...	Estate	90	20.0	California	Dry Creek Valley	Sonc
65497	US	This wine wears its 15.8% alcohol	Block 24	90	31.0	California	Napa Valley	Na

	country	description	designation	points	price	province	region_1	regior
		better than ...						
65498	Spain	A unique take on Manzanilla Sherry, which is o...	Manzanilla	90	10.0	Andalucia	Jerez	N

65499 rows × 15 columns

Your turn

If you haven't started the exercise, you can start now.