

# Data Analysis with Pandas

April 16, 2019

## 1 Pandas

Pandas is a python package that deals mostly with : - Series (1 dimensional homogenous array) - DataFrame (2 Dimensional heterogenous array) - Panel (general 3D array)

### 1.1 Outline

1. Creating, Reading and Writing Data
2. Indexing, Selecting and Assigning
3. Summary Functions and Maps
4. Grouping
5. Data Types and Missing Data

```
In [1]: #import python libraries
import pandas as pd
import numpy as np
```

### 1.2 Series

A series is a 1 dimensional labelled array containing data of the same type. The axis labels are often referred to index ### Creating a Series

```
In [2]: #Create a Pandas Series from the 5 numbers generated randomly from the normal distribu
s1=pd.Series(np.random.randn(5))
```

```
In [3]: s1
```

```
Out[3]: 0    0.702841
1    0.202842
2   -0.631774
3    0.656059
4   -0.081952
dtype: float64
```

```
In [4]: # Optionally specify name and index attributes
s2=pd.Series([20,30,40,50],name="age",index=range(4))
```

```
In [5]: s3=pd.Series(["dan","liya","makena","kendi"],name="family",index=["a","b","c","d"])
```

```
In [6]: s3
```

```
Out[6]: a      dan  
       b      liya  
       c      makena  
       d      kendi  
       Name: family, dtype: object
```

```
In [7]: s1.index
```

```
Out[7]: RangeIndex(start=0, stop=5, step=1)
```

```
In [8]: s2.index
```

```
Out[8]: RangeIndex(start=0, stop=4, step=1)
```

```
In [9]: s3.index
```

```
Out[9]: Index(['a', 'b', 'c', 'd'], dtype='object')
```

```
In [10]: mydict={"x":4,"e":2}  
        s4=pd.Series(mydict)
```

```
In [11]: s4.index
```

```
Out[11]: Index(['e', 'x'], dtype='object')
```

### 1.2.1 Accessing Elements of a Series

You can access individual elements of a Series Using an index

```
In [12]: s4[0]
```

```
Out[12]: 2
```

```
In [13]: s4
```

```
Out[13]: e      2  
        x      4  
        dtype: int64
```

```
In [14]: s4['e']
```

```
Out[14]: 2
```

Many ways to Slice Pandas Series

```
In [15]: s1
```

```
Out[15]: 0    0.702841
         1    0.202842
         2   -0.631774
         3    0.656059
         4   -0.081952
         dtype: float64
```

```
In [16]: s1[1:4] #returns elements from index 1 to index 3
```

```
Out[16]: 1    0.202842
         2   -0.631774
         3    0.656059
         dtype: float64
```

```
In [17]: s1[[1,3,2]] #return elements at those indexes in the order specified  
         #notice the order in which they are returned
```

```
Out[17]: 1    0.202842
         3    0.656059
         2   -0.631774
         dtype: float64
```

```
In [18]: s1.mean()
```

```
Out[18]: 0.16960316816352594
```

```
In [19]: s1.median()
```

```
Out[19]: 0.20284166135976278
```

- You can select part of the series based on the values

```
In [20]: s1[s1>0] #All positive values
```

```
Out[20]: 0    0.702841
         1    0.202842
         3    0.656059
         dtype: float64
```

```
In [21]: s1[s1>s1.mean()] #Values above the mean
```

```
Out[21]: 0    0.702841
         1    0.202842
         3    0.656059
         dtype: float64
```

- You can pass Series data types to numpy functions to perform vector operations

```
In [22]: np.mean(s1)
```

```
Out[22]: 0.16960316816352594
```

```
In [23]: s1
```

```
Out[23]: 0    0.702841  
         1    0.202842  
         2   -0.631774  
         3    0.656059  
         4   -0.081952  
         dtype: float64
```

```
In [24]: s1=s1*2
```

```
In [25]: s1
```

```
Out[25]: 0    1.405683  
         1    0.405683  
         2   -1.263549  
         3    1.312118  
         4   -0.163904  
         dtype: float64
```

## 1.2.2 Popular Attributes and Methods:

Attribute/Method	Description
dtype	data type of values in series
empty	True if series is empty
size	number of elements
values	Returns values as ndarray
head()	First n elements
tail()	Last n elements

**\*\* Exercise 1** Create a Series of 30 random values from the standard normal distribution and explore the above attributes and methods\*\*

## 1.3 Pandas DataFrame

Pandas DataFrame is two-dimensional, size-mutable, heterogeneous tabular data structure with labeled rows and columns(axes). Can be thought of a dictionary-like container to store python Series objects.

### 1.3.1 Creating a DataFrame

From a dictionary

```
In [26]: data={"Apples":30,"Bananas":21}
```

```
In [27]: fruits=pd.DataFrame(data,index=[0])
```

```
In [28]: fruits
```

```

Out[28]:    Apples  Bananas
         0      30      21

In [29]: data={"Apples":pd.Series([25,21,30]),"Bananas":pd.Series([25,3,2])}

In [30]: more_fruits=pd.DataFrame(data)

In [31]: more_fruits

Out[31]:    Apples  Bananas
         0      25      25
         1      21       3
         2      30       2

In [32]: data={"Apples":[25,21,30],"Bananas":[25,3,2]}

In [33]: other_fruits=pd.DataFrame(data, index=["One","Two","Three"])

In [34]: other_fruits

Out[34]:    Apples  Bananas
One      25      25
Two      21       3
Three    30       2

In [35]: more_fruits

Out[35]:    Apples  Bananas
         0      25      25
         1      21       3
         2      30       2

In [36]: more_fruits["Pineapples"]=0

In [37]: more_fruits

Out[37]:    Apples  Bananas  Pineapples
         0      25      25           0
         1      21       3           0
         2      30       2           0

In [38]: more_fruits["Orange"]=[2,5,3]

In [39]: more_fruits

Out[39]:    Apples  Bananas  Pineapples  Orange
         0      25      25           0       2
         1      21       3           0       5
         2      30       2           0       3

In [40]: more_fruits["Apples"]=[7,7,7]

```

```

In [41]: more_fruits

Out[41]:
   Apples  Bananas  Pineapples  Orange
0        7        25           0        2
1        7         3           0        5
2        7         2           0        3

In [42]: #Write the data to a csv file
more_fruits.to_csv("fruits.csv")

In [43]: #Read data from a csv file
reviews=pd.read_csv("wine reviews.csv",index_col=0)

In [44]: reviews.columns

Out[44]: Index(['country', 'description', 'designation', 'points', 'price', 'province',
               'region_1', 'region_2', 'taster_name', 'taster_twitter_handle', 'title',
               'variety', 'winery'],
              dtype='object')

In [45]: reviews.index

Out[45]: Int64Index([      0,      1,      2,      3,      4,      5,      6,      7,
                    8,      9,
                    ...,
                    129961, 129962, 129963, 129964, 129965, 129966, 129967, 129968,
                    129969, 129970],
                  dtype='int64', length=129971)

In [46]: reviews.shape

Out[46]: (129971, 13)

In [47]: reviews.dtypes

Out[47]: country                object
description                    object
designation                    object
points                        int64
price                        float64
province                      object
region_1                      object
region_2                      object
taster_name                   object
taster_twitter_handle         object
title                        object
variety                      object
winery                       object
dtype: object

```

### 1.3.2 Data Frame Attributes

df.attribute	description
dtypes	list the types of the columns
columns	list the column names
axes	list the row labels and column names
ndim	number of dimensions
size	number of elements
shape	return a tuple representing the dimensionality
values	numpy representation of the data

In [48]: reviews.axes

```
Out [48]: [Int64Index([      0,      1,      2,      3,      4,      5,      6,      7,
                    8,      9,
                    ...
                    129961, 129962, 129963, 129964, 129965, 129966, 129967, 129968,
                    129969, 129970],
                    dtype='int64', length=129971),
          Index(['country', 'description', 'designation', 'points', 'price', 'province',
                'region_1', 'region_2', 'taster_name', 'taster_twitter_handle', 'title',
                'variety', 'winery'],
                dtype='object')]
```

### 1.3.3 Data Frame Methods

Unlike attributes, python methods have paranthesis.

All attributes and methods can be listed with a dir() function: dir(data)

df.method	description
head([n]),tail([n])	first/last n rows
describe()	generate descriptive statistics (for numeric columns only)
max(), min()	return max/ min values for all numeric columns
mean(),median()	return mean/median values for all numeric columns
std()	standard deviation
sample([n])	returns a random sample of the data frame
dropna()	drop all the records with missing values

In [49]: reviews.head()

```
Out [49]:
```

	country	description \
0	Italy	Aromas include tropical fruit, broom, brimston...
1	Portugal	This is ripe and fruity, a wine that is smooth...
2	US	Tart and snappy, the flavors of lime flesh and...
3	US	Pineapple rind, lemon pith and orange blossom ...
4	US	Much like the regular bottling from 2012, this...

  

	designation	points	price	province \
0	Vulkà Bianco	87	NaN	Sicily & Sardinia

1		Avidagos	87	15.0	Douro
2		NaN	87	14.0	Oregon
3		Reserve Late Harvest	87	13.0	Michigan
4	Vintner's Reserve Wild Child Block		87	65.0	Oregon

	region_1	region_2	taster_name \
0	Etna	NaN	Kerin OKeefe
1	NaN	NaN	Roger Voss
2	Willamette Valley	Willamette Valley	Paul Gregutt
3	Lake Michigan Shore	NaN	Alexander Peartree
4	Willamette Valley	Willamette Valley	Paul Gregutt

	taster_twitter_handle	title \
0	@kerinokeefe	Nicosia 2013 Vulkà Bianco (Etna)
1	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)
2	@paulgwineã	Rainstorm 2013 Pinot Gris (Willamette Valley)
3	NaN	St. Julian 2013 Reserve Late Harvest Riesling ...
4	@paulgwineã	Sweet Cheeks 2012 Vintner's Reserve Wild Child...

	variety	winery
0	White Blend	Nicosia
1	Portuguese Red	Quinta dos Avidagos
2	Pinot Gris	Rainstorm
3	Riesling	St. Julian
4	Pinot Noir	Sweet Cheeks

In [50]: reviews.head(10)

Out [50]:

	country	description \
0	Italy	Aromas include tropical fruit, broom, brimston...
1	Portugal	This is ripe and fruity, a wine that is smooth...
2	US	Tart and snappy, the flavors of lime flesh and...
3	US	Pineapple rind, lemon pith and orange blossom ...
4	US	Much like the regular bottling from 2012, this...
5	Spain	Blackberry and raspberry aromas show a typical...
6	Italy	Here's a bright, informal red that opens with ...
7	France	This dry and restrained wine offers spice in p...
8	Germany	Savory dried thyme notes accent sunnier flavor...
9	France	This has great depth of flavor with its fresh ...

	designation	points	price	province \
0	Vulkà Bianco	87	NaN	Sicily & Sardinia
1	Avidagos	87	15.0	Douro
2	NaN	87	14.0	Oregon
3	Reserve Late Harvest	87	13.0	Michigan
4	Vintner's Reserve Wild Child Block	87	65.0	Oregon
5	Ars In Vitro	87	15.0	Northern Spain
6	Belsito	87	16.0	Sicily & Sardinia



7		NaN	87	24.0	Alsace
8		Shine	87	12.0	Rheinhessen
9		Les Natures	87	27.0	Alsace

	region_1	region_2	taster_name \
0	Etna	NaN	Kerin OKeefe
1	NaN	NaN	Roger Voss
2	Willamette Valley	Willamette Valley	Paul Gregutt
3	Lake Michigan Shore	NaN	Alexander Peartree
4	Willamette Valley	Willamette Valley	Paul Gregutt
5	Navarra	NaN	Michael Schachner
6	Vittoria	NaN	Kerin OKeefe
7	Alsace	NaN	Roger Voss
8	NaN	NaN	Anna Lee C. Iijima
9	Alsace	NaN	Roger Voss

	taster_twitter_handle	title \
0	@kerinokeefe	Nicosia 2013 Vulkà Bianco (Etna)
1	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)
2	@paulgwineã	Rainstorm 2013 Pinot Gris (Willamette Valley)
3	NaN	St. Julian 2013 Reserve Late Harvest Riesling ...
4	@paulgwineã	Sweet Cheeks 2012 Vintner's Reserve Wild Child...
5	@wineschach	Tandem 2011 Ars In Vitro Tempranillo-Merlot (N...
6	@kerinokeefe	Terre di Giurfo 2013 Belsito Frappato (Vittoria)
7	@vossroger	Trimbach 2012 Gewurztraminer (Alsace)
8	NaN	Heinz Eifel 2013 Shine Gewürztraminer (Rheinhe...
9	@vossroger	Jean-Baptiste Adam 2012 Les Natures Pinot Gris...

	variety	winery
0	White Blend	Nicosia
1	Portuguese Red	Quinta dos Avidagos
2	Pinot Gris	Rainstorm
3	Riesling	St. Julian
4	Pinot Noir	Sweet Cheeks
5	Tempranillo-Merlot	Tandem
6	Frappato	Terre di Giurfo
7	Gewürztraminer	Trimbach
8	Gewürztraminer	Heinz Eifel
9	Pinot Gris	Jean-Baptiste Adam

In [51]: reviews.describe()

```
Out[51]:
```

	points	price
count	129971.000000	120975.000000
mean	88.447138	35.363389
std	3.039730	41.022218
min	80.000000	4.000000
25%	86.000000	17.000000

50%	88.000000	25.000000
75%	91.000000	42.000000
max	100.000000	3300.000000

### 1.3.4 Data Slicing

**\*\* Selecting Using Columns\*\***

In [52]: reviews.columns

Out [52]: Index(['country', 'description', 'designation', 'points', 'price', 'province',  
'region\_1', 'region\_2', 'taster\_name', 'taster\_twitter\_handle', 'title',  
'variety', 'winery'],  
dtype='object')

In [53]: reviews.country

Out [53]: 0 Italy  
1 Portugal  
2 US  
3 US  
4 US  
5 Spain  
6 Italy  
7 France  
8 Germany  
9 France  
10 US  
11 France  
12 US  
13 Italy  
14 US  
15 Germany  
16 Argentina  
17 Argentina  
18 Spain  
19 US  
20 US  
21 US  
22 Italy  
23 US  
24 Italy  
25 US  
26 Italy  
27 Italy  
28 Italy  
29 US  
...  
129941 US

```

129942      US
129943      Italy
129944      Israel
129945      US
129946      Germany
129947      Italy
129948      Argentina
129949      US
129950      US
129951      France
129952      US
129953      New Zealand
129954      New Zealand
129955      New Zealand
129956      New Zealand
129957      Spain
129958      New Zealand
129959      France
129960      Portugal
129961      Italy
129962      Italy
129963      Israel
129964      France
129965      France
129966      Germany
129967      US
129968      France
129969      France
129970      France
Name: country, Length: 129971, dtype: object

```

```
In [54]: reviews["description"];
```

```
In [55]: reviews[["country","description","designation"]]; #Select multiple columns
```

**\*\* Selecting rows\*\*** In addition to head and tail methods we call also use indexes the methods loc and iloc can also be used

```
In [56]: reviews[0:10];#First 10 rows
```

```
In [57]: reviews[-5:]; # Last 5 rows
```

```
In [58]: reviews[0:10:2];#Notice the step
```

```
In [59]: reviews.loc[100] #Item at the specific index
```

```

Out[59]: country      US
         description  Fresh apple, lemon and pear flavors are accent...
         designation  NaN

```

points	88
price	18
province	New York
region_1	Finger Lakes
region_2	Finger Lakes
taster_name	Anna Lee C. Iijima
taster_twitter_handle	NaN
title	Ventosa 2015 Pinot Gris (Finger Lakes)
variety	Pinot Gris
winery	Ventosa

Name: 100, dtype: object

In [60]: reviews.iloc[100]

country	US
description	Fresh apple, lemon and pear flavors are accent...
designation	NaN
points	88
price	18
province	New York
region_1	Finger Lakes
region_2	Finger Lakes
taster_name	Anna Lee C. Iijima
taster_twitter_handle	NaN
title	Ventosa 2015 Pinot Gris (Finger Lakes)
variety	Pinot Gris
winery	Ventosa

Name: 100, dtype: object

In [61]: reviews.iloc[:5]

country	description \
0 Italy	Aromas include tropical fruit, broom, brimston...
1 Portugal	This is ripe and fruity, a wine that is smooth...
2 US	Tart and snappy, the flavors of lime flesh and...
3 US	Pineapple rind, lemon pith and orange blossom ...
4 US	Much like the regular bottling from 2012, this...

  

designation	points	price	province \
0 Vulkà Bianco	87	NaN	Sicily & Sardinia
1 Avidagos	87	15.0	Douro
2 NaN	87	14.0	Oregon
3 Reserve Late Harvest	87	13.0	Michigan
4 Vintner's Reserve Wild Child Block	87	65.0	Oregon

  

region_1	region_2	taster_name \
0 Etna	NaN	Kerin OKeefe
1 NaN	NaN	Roger Voss
2 Willamette Valley	Willamette Valley	Paul Gregutt

3	Lake Michigan Shore	NaN	Alexander Peartree
4	Willamette Valley	Willamette Valley	Paul Gregutt

	taster_twitter_handle		title \
0	@kerinokeefe	Nicosia 2013 Vulkà Bianco	(Etna)
1	@vossroger	Quinta dos Avidagos 2011 Avidagos Red	(Douro)
2	@paulgwineã	Rainstorm 2013 Pinot Gris	(Willamette Valley)
3	NaN	St. Julian 2013 Reserve Late Harvest Riesling	...
4	@paulgwineã	Sweet Cheeks 2012 Vintner's Reserve Wild Child...	

	variety	winery
0	White Blend	Nicosia
1	Portuguese Red	Quinta dos Avidagos
2	Pinot Gris	Rainstorm
3	Riesling	St. Julian
4	Pinot Noir	Sweet Cheeks

In [62]: reviews.loc[:5]

Out[62]:

	country	description \
0	Italy	Aromas include tropical fruit, broom, brimston...
1	Portugal	This is ripe and fruity, a wine that is smooth...
2	US	Tart and snappy, the flavors of lime flesh and...
3	US	Pineapple rind, lemon pith and orange blossom ...
4	US	Much like the regular bottling from 2012, this...
5	Spain	Blackberry and raspberry aromas show a typical...

	designation	points	price	province \
0	Vulkà Bianco	87	NaN	Sicily & Sardinia
1	Avidagos	87	15.0	Douro
2	NaN	87	14.0	Oregon
3	Reserve Late Harvest	87	13.0	Michigan
4	Vintner's Reserve Wild Child Block	87	65.0	Oregon
5	Ars In Vitro	87	15.0	Northern Spain

	region_1	region_2	taster_name \
0	Etna	NaN	Kerin OKeefe
1	NaN	NaN	Roger Voss
2	Willamette Valley	Willamette Valley	Paul Gregutt
3	Lake Michigan Shore	NaN	Alexander Peartree
4	Willamette Valley	Willamette Valley	Paul Gregutt
5	Navarra	NaN	Michael Schachner

	taster_twitter_handle		title \
0	@kerinokeefe	Nicosia 2013 Vulkà Bianco	(Etna)
1	@vossroger	Quinta dos Avidagos 2011 Avidagos Red	(Douro)
2	@paulgwineã	Rainstorm 2013 Pinot Gris	(Willamette Valley)
3	NaN	St. Julian 2013 Reserve Late Harvest Riesling	...

```

4      @paulgwineã Sweet Cheeks 2012 Vintner's Reserve Wild Child...
5      @wineschach Tandem 2011 Ars In Vitro Tempranillo-Merlot (N...

```

```

          variety          winery
0      White Blend      Nicosia
1      Portuguese Red  Quinta dos Avidagos
2      Pinot Gris      Rainstorm
3      Riesling        St. Julian
4      Pinot Noir      Sweet Cheeks
5      Tempranillo-Merlot      Tandem

```

```
In [63]: reviews.country.loc[:5] #First 5 items from the column country
```

```

Out [63]: 0      Italy
          1      Portugal
          2      US
          3      US
          4      US
          5      Spain
          Name: country, dtype: object

```

```
In [64]: reviews[["description", "country"]].loc[:5]
```

```

Out [64]:          description  country
0  Aromas include tropical fruit, broom, brimston...  Italy
1  This is ripe and fruity, a wine that is smooth... Portugal
2  Tart and snappy, the flavors of lime flesh and...    US
3  Pineapple rind, lemon pith and orange blossom ...    US
4  Much like the regular bottling from 2012, this...    US
5  Blackberry and raspberry aromas show a typical...  Spain

```

```
In [65]: reviews.iloc[[1,2,3,4,5]] #specific index
```

```

Out [65]:    country          description \
1  Portugal  This is ripe and fruity, a wine that is smooth...
2      US  Tart and snappy, the flavors of lime flesh and...
3      US  Pineapple rind, lemon pith and orange blossom ...
4      US  Much like the regular bottling from 2012, this...
5  Spain  Blackberry and raspberry aromas show a typical...

```

```

          designation  points  price  province \
1          Avidagos      87   15.0      Douro
2          NaN      87   14.0      Oregon
3      Reserve Late Harvest      87   13.0      Michigan
4  Vintner's Reserve Wild Child Block      87   65.0      Oregon
5          Ars In Vitro      87   15.0  Northern Spain

```

```

          region_1      region_2  taster_name \
1          NaN      NaN      Roger Voss

```

2	Willamette Valley	Willamette Valley	Paul Gregutt
3	Lake Michigan Shore	NaN	Alexander Peartree
4	Willamette Valley	Willamette Valley	Paul Gregutt
5	Navarra	NaN	Michael Schachner

	taster_twitter_handle	title \
1	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)
2	@paulgwineã	Rainstorm 2013 Pinot Gris (Willamette Valley)
3	NaN	St. Julian 2013 Reserve Late Harvest Riesling ...
4	@paulgwineã	Sweet Cheeks 2012 Vintner's Reserve Wild Child...
5	@wineschach	Tandem 2011 Ars In Vitro Tempranillo-Merlot (N...

	variety	winery
1	Portuguese Red	Quinta dos Avidagos
2	Pinot Gris	Rainstorm
3	Riesling	St. Julian
4	Pinot Noir	Sweet Cheeks
5	Tempranillo-Merlot	Tandem

In [66]: reviews.loc[[1,2,3,4,5]]

Out [66]:

	country	description \
1	Portugal	This is ripe and fruity, a wine that is smooth...
2	US	Tart and snappy, the flavors of lime flesh and...
3	US	Pineapple rind, lemon pith and orange blossom ...
4	US	Much like the regular bottling from 2012, this...
5	Spain	Blackberry and raspberry aromas show a typical...

	designation	points	price	province \
1	Avidagos	87	15.0	Douro
2	NaN	87	14.0	Oregon
3	Reserve Late Harvest	87	13.0	Michigan
4	Vintner's Reserve Wild Child Block	87	65.0	Oregon
5	Ars In Vitro	87	15.0	Northern Spain

	region_1	region_2	taster_name \
1	NaN	NaN	Roger Voss
2	Willamette Valley	Willamette Valley	Paul Gregutt
3	Lake Michigan Shore	NaN	Alexander Peartree
4	Willamette Valley	Willamette Valley	Paul Gregutt
5	Navarra	NaN	Michael Schachner

	taster_twitter_handle	title \
1	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)
2	@paulgwineã	Rainstorm 2013 Pinot Gris (Willamette Valley)
3	NaN	St. Julian 2013 Reserve Late Harvest Riesling ...
4	@paulgwineã	Sweet Cheeks 2012 Vintner's Reserve Wild Child...
5	@wineschach	Tandem 2011 Ars In Vitro Tempranillo-Merlot (N...

	variety	winery
1	Portuguese Red	Quinta dos Avidagos
2	Pinot Gris	Rainstorm
3	Riesling	St. Julian
4	Pinot Noir	Sweet Cheeks
5	Tempranillo-Merlot	Tandem

```
In [67]: reviews[["country","province","region_1","region_2"]].iloc[[0,1,10,100]] #Select columns
```

```
Out[67]:
```

	country	province	region_1	region_2
0	Italy	Sicily & Sardinia	Etna	NaN
1	Portugal	Douro	NaN	NaN
10	US	California	Napa Valley	Napa
100	US	New York	Finger Lakes	Finger Lakes

```
In [68]: reviews.loc[[0,1,10,100],["country","province","region_1","region_2"]] #another way of
```

```
Out[68]:
```

	country	province	region_1	region_2
0	Italy	Sicily & Sardinia	Etna	NaN
1	Portugal	Douro	NaN	NaN
10	US	California	Napa Valley	Napa
100	US	New York	Finger Lakes	Finger Lakes

```
In [69]: reviews.iloc[[0,1,10,100],["country","province","region_1","region_2"]]; #iloc only t
```

-----

TypeError

Traceback (most recent call last)

```
<ipython-input-69-64cfd9994f3f> in <module>()
----> 1 reviews.iloc[[0,1,10,100],["country","province","region_1","region_2"]]; #iloc only t
```

```
~\Anaconda3\lib\site-packages\pandas\core\indexing.py in __getitem__(self, key)
1323         except (KeyError, IndexError):
1324             pass
-> 1325         return self._getitem_tuple(key)
1326     else:
1327         key = com._apply_if_callable(key, self.obj)
```

```
~\Anaconda3\lib\site-packages\pandas\core\indexing.py in _getitem_tuple(self, tup)
1660     def _getitem_tuple(self, tup):
1661
-> 1662         self._has_valid_tuple(tup)
1663         try:
1664             return self._getitem_lowerdim(tup)
```



```

~\Anaconda3\lib\site-packages\pandas\core\indexing.py in _has_valid_tuple(self, key)
187         if i >= self.obj.ndim:
188             raise IndexingError('Too many indexers')
--> 189         if not self._has_valid_type(k, i):
190             raise ValueError("Location based indexing can only have [%s] "
191                               "types" % self._valid_types)

~\Anaconda3\lib\site-packages\pandas\core\indexing.py in _has_valid_type(self, key, axis)
1597         return self._is_valid_integer(key, axis)
1598         elif is_list_like_indexer(key):
-> 1599             return self._is_valid_list_like(key, axis)
1600         return False
1601

~\Anaconda3\lib\site-packages\pandas\core\indexing.py in _is_valid_list_like(self, key, axis)
1653         l = len(ax)
1654         if (hasattr(arr, '__len__') and len(arr) and
-> 1655             (arr.max() >= 1 or arr.min() < -1)):
1656             raise IndexError("positional indexers are out-of-bounds")
1657

~\Anaconda3\lib\site-packages\numpy\core\_methods.py in _amax(a, axis, out, keepdims)
24 # small reductions
25 def _amax(a, axis=None, out=None, keepdims=False):
---> 26     return umr_maximum(a, axis, None, out, keepdims)
27
28 def _amin(a, axis=None, out=None, keepdims=False):

```

TypeError: cannot perform reduce with flexible type

In [70]: reviews.columns

Out[70]: Index(['country', 'description', 'designation', 'points', 'price', 'province',  
'region\_1', 'region\_2', 'taster\_name', 'taster\_twitter\_handle', 'title',  
'variety', 'winery'],  
dtype='object')

In [71]: reviews.iloc[[0,1,10,100],[0,5,6,7]] *#You would have to know the indexes of the columns*

Out[71]:

	country	province	region_1	region_2
0	Italy	Sicily & Sardinia	Etna	NaN
1	Portugal	Douro	NaN	NaN

10	US	California	Napa Valley	Napa
100	US	New York	Finger Lakes	Finger Lakes

**\*\* Exercise: Extract the first 100 rows of the columns country and price - Use loc - Use iloc - Use index slices**

**\*\***

```
In [72]: reviews.loc[0:100,["country","price"]];
```

```
In [73]: reviews.columns
```

```
Out[73]: Index(['country', 'description', 'designation', 'points', 'price', 'province',
               'region_1', 'region_2', 'taster_name', 'taster_twitter_handle', 'title',
               'variety', 'winery'],
              dtype='object')
```

```
In [74]: reviews.iloc[0:100,[0,4]];
```

```
In [75]: reviews[["country","price"]][0:100];
```

**Select data depending on some values e.g. Wines from Italy**

```
In [76]: reviews.head()
```

```
Out[76]:
```

	country	description \
0	Italy	Aromas include tropical fruit, broom, brimston...
1	Portugal	This is ripe and fruity, a wine that is smooth...
2	US	Tart and snappy, the flavors of lime flesh and...
3	US	Pineapple rind, lemon pith and orange blossom ...
4	US	Much like the regular bottling from 2012, this...

  

	designation	points	price	province \
0	Vulkà Bianco	87	NaN	Sicily & Sardinia
1	Avidagos	87	15.0	Douro
2	NaN	87	14.0	Oregon
3	Reserve Late Harvest	87	13.0	Michigan
4	Vintner's Reserve Wild Child Block	87	65.0	Oregon

  

	region_1	region_2	taster_name \
0	Etna	NaN	Kerin OKeefe
1	NaN	NaN	Roger Voss
2	Willamette Valley	Willamette Valley	Paul Gregutt
3	Lake Michigan Shore	NaN	Alexander Peartree
4	Willamette Valley	Willamette Valley	Paul Gregutt

  

	taster_twitter_handle	title \
0	@kerinokeefe	Nicosia 2013 Vulkà Bianco (Etna)
1	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)
2	@paulgwineã	Rainstorm 2013 Pinot Gris (Willamette Valley)

```

3          NaN  St. Julian 2013 Reserve Late Harvest Riesling ...
4    @paulgwineã Sweet Cheeks 2012 Vintner's Reserve Wild Child...

```

```

          variety          winery
0    White Blend      Nicosia
1 Portuguese Red  Quinta dos Avidagos
2    Pinot Gris      Rainstorm
3    Riesling        St. Julian
4    Pinot Noir      Sweet Cheeks

```

```
In [77]: reviews[reviews.country=="Italy"]; #Wines from Italy
```

```
In [78]: reviews.loc[reviews.country=="Italy"]; #Wines from Italy
```

```
In [79]: reviews.loc[reviews.country=="Italy",["price","winery"]]; #Wineries in Italy and thei
```

```
In [80]: #Extract all reviews with at least 95 points from Australia or New Zealand
reviews.loc[(reviews.points>=95) & (reviews.country.isin(["Australia","New Zealand"])]
```

```

Out[80]:          country          description \
345    Australia  This wine contains some material over 100 year...
346    Australia  This deep brown wine smells like a damp, mossy...
348    Australia  Deep mahogany. Dried fig and black tea on the ...
349    Australia  RunRig is always complex, and the 2012 doesn't...
356    Australia  Dusty, firm, powerful: just a few apt descript...
360    Australia  Bacon and tapenade elements merge easily on th...
365    Australia  The Taylor family selected Clare Valley for it...
14354   Australia  This wine's concentrated dark fruit shows in t...
16538   Australia  Rich, dense and intense, this is a big, muscul...
28573   Australia  Astralis has become one of Australia's top col...
34502   Australia  This prodigious wine showcases Barossa's abili...
34506   Australia  If Standish's Relic is the feminine side of Sh...
38988   Australia  Penfolds Bin 707 has leapt in quality over the...
39059   Australia  The Taylor family selected Clare Valley for it...
39961   Australia  As unevolved as they are, the dense and multil...
39962   Australia  Seamless luxury from stem to stern, this baby...
45809   Australia  The 2007 Astralis impresses for its combinatio...
56953   Australia  This inky, embryonic wine deserves to be cella...
56956   Australia  You may have to scour the country to secure so...
56957   Australia  Thorn Clarke has taken its Shiraz to a new lev...
56959   Australia  Is this the Yin to Grange's Yang? The wines ar...
59977   Australia  This is a top example of the classic Australia...
59984   Australia  This is a throwback to those brash, flavor-exu...
67096   Australia  Just a tiny serving of this dark nectar will l...
67101   Australia  This Muscat is the color of dark coffee, with ...
76392   Australia  When the alcohol levels are reined in to appro...
77028   Australia  This has all the size and weight you've come t...
77036   Australia  RWT (unromantically derived from Red Wine Tri...
77037   Australia  Winemaker Dave Powell is no longer with Torbre...

```

77042	Australia	This is likely the most ageworthy Shiraz winem...
77044	Australia	The fruit for this offering comes from the Gre...
77046	Australia	With aromas and flavors that range widely from...
83357	Australia	A throwback to the monster Shiraz style of old...
84815	Australia	The Factor is always one of Torbreck's biggest...
84816	Australia	Nashwauk is Kaesler's McLaren Vale project, fi...
87128	Australia	This full-bodied, muscular Shiraz is built for...
87137	Australia	Perhaps the best young wine I've tasted from M...
87143	Australia	This is wonderfully complex and aromatic, with...
91851	New Zealand	This full-bodied, richly tannic wine delivers...
98386	Australia	One of the more approachable of the d'Arenberg...
99318	Australia	From vines planted in 1912, this has been an i...
99330	Australia	This Cabernet equivalent to Grange has explode...
99340	Australia	This rich, opulent wine carries its massive oa...
109427	Australia	This wine is dark brown in hue with a greenish...
109434	Australia	D'Arenberg's lineup of single-vineyard Shiraze...
122421	Australia	Despite this wine's weight and richness, it re...
122430	Australia	These blends are traditional in Australiathey...
122507	New Zealand	This blend of Cabernet Sauvignon (62.5%), Merl...
122939	Australia	Full-bodied and plush yet vibrant and imbued w...

		designation	points	price \
345		Rare	100	350.0
346		Rare	98	350.0
348		Grand	97	100.0
349		RunRig	97	225.0
356		Georgia's Paddock	95	85.0
360		Descendant	95	125.0
365	St. Andrews Single Vineyard Release		95	60.0
14354		Old Vine	95	60.0
16538		The Family Tree	95	65.0
28573		Astralis	95	350.0
34502		The Relic	98	135.0
34506	The Standish Single Vineyard		96	135.0
38988		Bin 707	95	200.0
39059	St. Andrews Single Vineyard Release		95	60.0
39961		Grange	96	185.0
39962		RWT	95	70.0
45809		Astralis	95	225.0
56953		Grange	99	850.0
56956		Andelmonde	97	95.0
56957	Ron Thorn Single Vineyard		96	89.0
56959		Hill of Grace	96	820.0
59977		The Peake	96	150.0
59984		One	95	95.0
67096		Calliope Rare	98	86.0
67101		Rare	95	300.0
76392		Georgia's Paddock	95	85.0

77028		Grange	98	850.0
77036		RWT	96	150.0
77037		RunRig	96	225.0
77042		Eligo	95	100.0
77044		R Reserve	95	105.0
77046		The Factor	95	125.0
83357		Grange	96	500.0
84815		The Factor	95	125.0
84816		Beacon	95	145.0
87128		Amery Vineyard Block 6	96	120.0
87137		NaN	95	84.0
87143	St. Andrews Single Vineyard Release		95	60.0
91851		Homage	95	100.0
98386	Little Venice Single Vineyard		95	85.0
99318	Mount Edelstone Vineyard		95	200.0
99330		Bin 707	95	500.0
99340		Les Amis	95	185.0
109427		Rare	99	300.0
109434	The Swinging Malaysian Single Vineyard		96	85.0
122421		Amon-Ra Unfiltered	96	110.0
122430		Anaperenna	95	80.0
122507	SQM Gimblett Gravels Cabernets/Merlot		95	79.0
122939		The Factor	98	125.0

	province	region_1	region_2	taster_name \
345	Victoria	Rutherglen	NaN	Joe Czerwinski
346	Victoria	Rutherglen	NaN	Joe Czerwinski
348	Victoria	Rutherglen	NaN	Joe Czerwinski
349	South Australia	Barossa	NaN	Joe Czerwinski
356	Victoria	Heathcote	NaN	Joe Czerwinski
360	South Australia	Barossa Valley	NaN	Joe Czerwinski
365	South Australia	Clare Valley	NaN	Joe Czerwinski
14354	South Australia	Barossa Valley	NaN	Joe Czerwinski
16538	South Australia	Barossa Valley	NaN	Joe Czerwinski
28573	South Australia	Clarendon	NaN	Joe Czerwinski
34502	South Australia	Barossa Valley	NaN	Joe Czerwinski
34506	South Australia	Barossa Valley	NaN	Joe Czerwinski
38988	South Australia	South Australia	NaN	Joe Czerwinski
39059	South Australia	Clare Valley	NaN	Joe Czerwinski
39961	South Australia	South Australia	NaN	NaN
39962	South Australia	Barossa Valley	NaN	NaN
45809	South Australia	Clarendon	NaN	Joe Czerwinski
56953	South Australia	South Australia	NaN	Joe Czerwinski
56956	South Australia	Barossa Valley	NaN	Joe Czerwinski
56957	South Australia	Barossa	NaN	Joe Czerwinski
56959	South Australia	Eden Valley	NaN	Joe Czerwinski
59977	South Australia	McLaren Vale	NaN	Joe Czerwinski
59984	South Australia	Langhorne Creek	NaN	Joe Czerwinski

67096	Victoria	Rutherglen	NaN	Joe Czerwinski
67101	Victoria	Rutherglen	NaN	Joe Czerwinski
76392	Victoria	Heathcote	NaN	Joe Czerwinski
77028	South Australia	South Australia	NaN	Joe Czerwinski
77036	South Australia	Barossa Valley	NaN	Joe Czerwinski
77037	South Australia	Barossa Valley	NaN	Joe Czerwinski
77042	South Australia	Barossa	NaN	Joe Czerwinski
77044	South Australia	Barossa Valley	NaN	Joe Czerwinski
77046	South Australia	Barossa Valley	NaN	Joe Czerwinski
83357	South Australia	South Australia	NaN	Joe Czerwinski
84815	South Australia	Barossa	NaN	Joe Czerwinski
84816	South Australia	McLaren Vale	NaN	Joe Czerwinski
87128	South Australia	McLaren Vale	NaN	Joe Czerwinski
87137	Western Australia	Margaret River	NaN	Joe Czerwinski
87143	South Australia	Clare Valley	NaN	Joe Czerwinski
91851	Hawke's Bay	NaN	NaN	Joe Czerwinski
98386	South Australia	McLaren Vale	NaN	Joe Czerwinski
99318	South Australia	Eden Valley	NaN	Joe Czerwinski
99330	South Australia	South Australia	NaN	Joe Czerwinski
99340	South Australia	Barossa Valley	NaN	Joe Czerwinski
109427	Victoria	Rutherglen	NaN	Joe Czerwinski
109434	South Australia	McLaren Vale	NaN	Joe Czerwinski
122421	South Australia	Barossa Valley	NaN	Joe Czerwinski
122430	South Australia	Barossa Valley	NaN	Joe Czerwinski
122507	Hawke's Bay	NaN	NaN	Joe Czerwinski
122939	South Australia	Barossa Valley	NaN	Joe Czerwinski

	taster_twitter_handle \
345	@JoeCz
346	@JoeCz
348	@JoeCz
349	@JoeCz
356	@JoeCz
360	@JoeCz
365	@JoeCz
14354	@JoeCz
16538	@JoeCz
28573	@JoeCz
34502	@JoeCz
34506	@JoeCz
38988	@JoeCz
39059	@JoeCz
39961	NaN
39962	NaN
45809	@JoeCz
56953	@JoeCz
56956	@JoeCz
56957	@JoeCz

56959	@JoeCz
59977	@JoeCz
59984	@JoeCz
67096	@JoeCz
67101	@JoeCz
76392	@JoeCz
77028	@JoeCz
77036	@JoeCz
77037	@JoeCz
77042	@JoeCz
77044	@JoeCz
77046	@JoeCz
83357	@JoeCz
84815	@JoeCz
84816	@JoeCz
87128	@JoeCz
87137	@JoeCz
87143	@JoeCz
91851	@JoeCz
98386	@JoeCz
99318	@JoeCz
99330	@JoeCz
99340	@JoeCz
109427	@JoeCz
109434	@JoeCz
122421	@JoeCz
122430	@JoeCz
122507	@JoeCz
122939	@JoeCz

	title \
345	Chambers Rosewood Vineyards NV Rare Muscat (Ru...
346	Chambers Rosewood Vineyards NV Rare Muscadelle...
348	Chambers Rosewood Vineyards NV Grand Muscat (R...
349	Torbreck 2012 RunRig Shiraz-Viognier (Barossa)
356	Jasper Hill 2013 Georgia's Paddock Shiraz (Hea...
360	Torbreck 2012 Descendant Shiraz-Viognier (Baro...
365	Wakefield 2013 St. Andrews Single Vineyard Rel...
14354	Kaesler 2006 Old Vine Shiraz (Barossa Valley)
16538	Lambert 2013 The Family Tree Shiraz (Barossa V...
28573	Clarendon Hills 2005 Astralis Syrah (Clarendon)
34502	Standish 2006 The Relic Shiraz (Barossa Valley)
34506	Standish 2005 The Standish Single Vineyard Shi...
38988	Penfolds 2009 Bin 707 Cabernet Sauvignon (Sout...
39059	Wakefield 2013 St. Andrews Single Vineyard Rel...
39961	Penfolds 1996 Grange Shiraz (South Australia)
39962	Penfolds 1998 RWT Shiraz (Barossa Valley)
45809	Clarendon Hills 2007 Astralis Syrah (Clarendon)

56953 Penfolds 2010 Grange Shiraz (South Australia)  
 56956 Standish 2012 Andelmonde Shiraz (Barossa Valley)  
 56957 Thorn Clarke 2012 Ron Thorn Single Vineyard Sh...  
 56959 Henschke 2010 Hill of Grace Shiraz (Eden Valley)  
 59977 Hickinbotham 2013 The Peake Cabernet-Shiraz (M...  
 59984 Heartland 2013 One Red (Langhorne Creek)  
 67096 R.L. Buller & Son NV Calliope Rare Tokay (Ruth...  
 67101 Chambers Rosewood Vineyards NV Rare Muscat (Ru...  
 76392 Jasper Hill 2012 Georgia's Paddock Shiraz (Hea...  
 77028 Penfolds 2008 Grange Shiraz (South Australia)  
 77036 Penfolds 2009 RWT Shiraz (Barossa Valley)  
 77037 Torbreck 2007 RunRig Shiraz-Viognier (Barossa ...  
 77042 John Duval Wines 2010 Eligo Shiraz (Barossa)  
 77044 Kilikanoon 2009 R Reserve Shiraz (Barossa Valley)  
 77046 Torbreck 2007 The Factor Shiraz (Barossa Valley)  
 83357 Penfolds 2004 Grange Shiraz (South Australia)  
 84815 Torbreck 2012 The Factor Shiraz (Barossa)  
 84816 Nashwauk 2010 Beacon Shiraz (McLaren Vale)  
 87128 Kay Brothers 2012 Amery Vineyard Block 6 Shira...  
 87137 Moss Wood 2011 Cabernet Sauvignon (Margaret Ri...  
 87143 Wakefield 2012 St. Andrews Single Vineyard Rel...  
 91851 Trinity Hill 2013 Homage Syrah (Hawke's Bay)  
 98386 D'Arenberg 2010 Little Venice Single Vineyard ...  
 99318 Henschke 2014 Mount Edelstone Vineyard Shiraz ...  
 99330 Penfolds 2014 Bin 707 Cabernet Sauvignon (Sout...  
 99340 Torbreck 2013 Les Amis Grenache (Barossa Valley)  
 109427 Chambers Rosewood Vineyards NV Rare Muscadelle...  
 109434 D'Arenberg 2010 The Swinging Malaysian Single ...  
 122421 Glaetzer 2010 Amon-Ra Unfiltered Shiraz (Baros...  
 122430 Glaetzer 2010 Anaperenna Shiraz-Cabernet Sauvi...  
 122507 Squawking Magpie 2014 SQM Gimblett Gravels Cab...  
 122939 Torbreck 2013 The Factor Shiraz (Barossa Valley)

	variety	winery
345	Muscat	Chambers Rosewood Vineyards
346	Muscadelle	Chambers Rosewood Vineyards
348	Muscat	Chambers Rosewood Vineyards
349	Shiraz-Viognier	Torbreck
356	Shiraz	Jasper Hill
360	Shiraz-Viognier	Torbreck
365	Shiraz	Wakefield
14354	Shiraz	Kaesler
16538	Shiraz	Lambert
28573	Syrah	Clarendon Hills
34502	Shiraz	Standish
34506	Shiraz	Standish
38988	Cabernet Sauvignon	Penfolds
39059	Shiraz	Wakefield



39961	Shiraz	Penfolds
39962	Shiraz	Penfolds
45809	Syrah	Clarendon Hills
56953	Shiraz	Penfolds
56956	Shiraz	Standish
56957	Shiraz	Thorn Clarke
56959	Shiraz	Henschke
59977	Cabernet-Shiraz	Hickinbotham
59984	Red Blend	Heartland
67096	Tokay	R.L. Buller & Son
67101	Muscat	Chambers Rosewood Vineyards
76392	Shiraz	Jasper Hill
77028	Shiraz	Penfolds
77036	Shiraz	Penfolds
77037	Shiraz-Viognier	Torbreck
77042	Shiraz	John Duval Wines
77044	Shiraz	Kilikanoon
77046	Shiraz	Torbreck
83357	Shiraz	Penfolds
84815	Shiraz	Torbreck
84816	Shiraz	Nashwauk
87128	Shiraz	Kay Brothers
87137	Cabernet Sauvignon	Moss Wood
87143	Shiraz	Wakefield
91851	Syrah	Trinity Hill
98386	Shiraz	D'Arenberg
99318	Shiraz	Henschke
99330	Cabernet Sauvignon	Penfolds
99340	Grenache	Torbreck
109427	Muscadelle	Chambers Rosewood Vineyards
109434	Shiraz	D'Arenberg
122421	Shiraz	Glaetzer
122430	Shiraz-Cabernet Sauvignon	Glaetzer
122507	Bordeaux-style Red Blend	Squawking Magpie
122939	Shiraz	Torbreck

```
In [81]: reviews[reviews.country=="New Zealand"].loc[reviews.points>=95];
```

Exercise: 1. Select all wines with a price of more than 100 2. Select the designation of wines from South africa with a price of less than 100

### 1.3.5 Summary Functions and Maps

```
In [82]: reviews.describe() #Recall the use of describe in a dataframe
```

```
Out [82]:
```

	points	price
count	129971.000000	120975.000000
mean	88.447138	35.363389
std	3.039730	41.022218

min	80.000000	4.000000
25%	86.000000	17.000000
50%	88.000000	25.000000
75%	91.000000	42.000000
max	100.000000	3300.000000

```
In [83]: #To get median
reviews.points.median()
```

```
Out[83]: 88.0
```

```
In [96]: reviews.country.describe()
```

```
Out[96]: count      129908
unique         43
top            US
freq          54504
Name: country, dtype: object
```

```
In [84]: #What countries are represented in the dataset?
countries=reviews.country.unique()
```

```
In [85]: #How often does each country appear in the data set?
count_countries=reviews.country.value_counts()
```

```
In [86]: #Create a variable centered_price containing a version of the price column with the
centered_price=reviews.price-reviews.price.mean()
```

```
In [87]: centered_price;
```

```
In [90]: #Which wine is the best bargain? Has the highest points to price ratio

#find the ratio
ratio=reviews.points/reviews.price
```

```
In [91]: bargain=ratio.idxmax()
```

```
In [95]: reviews.loc[bargain,["title","country"]]
```

```
Out[95]: title      Bandit NV Merlot (California)
country              US
Name: 64590, dtype: object
```

Exercise: 1. How many tasters were asked to sample the wines 2. For each taster, how many wine samples did they review? 3. Combine the columns country and region\_1 to give you the country and region separated by a hyphen e.g. Italy-Etna

## Using map and lambda functions

```
In [97]: #Center the points on their mean
         points_mean=reviews.points.mean()
```

```
In [105]: reviews.points.map(lambda p:p-points_mean);
```

DataFrame.apply() method is used when you want to transform a whole dataframe by calling a custom method on each row

```
In [102]: def center_points(row):
         row.points=row.points-points_mean
         return row
```

```
In [106]: reviews.apply(center_points,axis="columns");
```

```
In [107]: reviews.country + '-' +reviews.region_1
```

```
Out[107]: 0          Italy-Etna
         1          NaN
         2      US-Willamette Valley
         3      US-Lake Michigan Shore
         4      US-Willamette Valley
         5      Spain-Navarra
         6      Italy-Vittoria
         7      France-Alsace
         8          NaN
         9      France-Alsace
        10      US-Napa Valley
        11      France-Alsace
        12      US-Alexander Valley
        13      Italy-Etna
        14      US-Central Coast
        15          NaN
        16      Argentina-Cafayate
        17      Argentina-Mendoza
        18      Spain-Ribera del Duero
        19      US-Virginia
        20      US-Virginia
        21      US-Oregon
        22      Italy-Sicilia
        23      US-Paso Robles
        24      Italy-Sicilia
        25      US-Sonoma Coast
        26      Italy-Terre Siciliane
        27      Italy-Terre Siciliane
        28      Italy-Cerasuolo di Vittoria
        29      US-Clarksburg
         ...
```

```

129941      US-Mendocino County
129942      US-Sonoma County
129943      Italy-Sicilia
129944      NaN
129945      US-Santa Ynez Valley
129946      NaN
129947      Italy-Terre Siciliane
129948      Argentina-San Juan
129949      US-Napa Valley
129950      US-Napa Valley
129951      France-Alsace
129952      US-Chiles Valley
129953      NaN
129954      NaN
129955      NaN
129956      NaN
129957      Spain-Rioja
129958      NaN
129959      France-Alsace
129960      NaN
129961      Italy-Sicilia
129962      Italy-Sicilia
129963      NaN
129964      France-Alsace
129965      France-Alsace
129966      NaN
129967      US-Oregon
129968      France-Alsace
129969      France-Alsace
129970      France-Alsace
Length: 129971, dtype: object

```

```

In [109]: #How many times does the word tropical appear in the descriptions?
reviews.description.map(lambda desc:"tropical" in desc).sum()

```

```

Out[109]: 3607

```

Try it by yourself: Create a Series showing how many times the words tropical and fruity appear in the descriptions

```

In [110]: def no_stars(data):
            if data["country"]=="Canada" or data["points"]>=95:
                return 3
            elif data["points"]>=85:
                return 2
            else:
                return 1

```

```

In [114]: reviews.apply(no_stars,axis=1);#1 stands for columns

```

### 1.3.6 Grouping and Sorting

Grouping can be used for many purposes.

In [118]: reviews

```
Out[118]:
```

	country	description \
0	Italy	Aromas include tropical fruit, broom, brimston...
1	Portugal	This is ripe and fruity, a wine that is smooth...
2	US	Tart and snappy, the flavors of lime flesh and...
3	US	Pineapple rind, lemon pith and orange blossom ...
4	US	Much like the regular bottling from 2012, this...
5	Spain	Blackberry and raspberry aromas show a typical...
6	Italy	Here's a bright, informal red that opens with ...
7	France	This dry and restrained wine offers spice in p...
8	Germany	Savory dried thyme notes accent sunnier flavor...
9	France	This has great depth of flavor with its fresh ...
10	US	Soft, supple plum envelopes an oaky structure ...
11	France	This is a dry wine, very spicy, with a tight, ...
12	US	Slightly reduced, this wine offers a chalky, t...
13	Italy	This is dominated by oak and oak-driven aromas...
14	US	Building on 150 years and six generations of w...
15	Germany	Zesty orange peels and apple notes abound in t...
16	Argentina	Baked plum, molasses, balsamic vinegar and che...
17	Argentina	Raw black-cherry aromas are direct and simple ...
18	Spain	Desiccated blackberry, leather, charred wood a...
19	US	Red fruit aromas pervade on the nose, with cig...
20	US	Ripe aromas of dark berries mingle with ample ...
21	US	A sleek mix of tart berry, stem and herb, alon...
22	Italy	Delicate aromas recall white flower and citrus...
23	US	This wine from the Geneseo district offers aro...
24	Italy	Aromas of prune, blackcurrant, toast and oak c...
25	US	Oak and earth intermingle around robust aromas...
26	Italy	Pretty aromas of yellow flower and stone fruit...
27	Italy	Aromas recall ripe dark berry, toast and a whi...
28	Italy	Aromas suggest mature berry, scorched earth, a...
29	US	Clarksburg is becoming a haven for Chenin Blan...
...	...	...
129941	US	A Chardonnay with an unusual companion, 8% Sém...
129942	US	This is classic in herbaceous aromas and flavo...
129943	Italy	A blend of Nero d'Avola and Syrah, this convey...
129944	Israel	Deep garnet in the glass, this has a nose of b...
129945	US	Hailing from one of the more popular vineyards...
129946	Germany	Plump, clingy peach and honey notes are cut wi...
129947	Italy	A blend of 65% Cabernet Sauvignon, 30% Merlot ...
129948	Argentina	Raspberry and cassis aromas are fresh and upri...
129949	US	There's no bones about the use of oak in this ...
129950	US	This opens with herbaceous dollops of thyme an...
129951	France	Hugely spicy this rich wine is described as sw...

129952	US	This Zinfandel from the eastern section of Nap...
129953	New Zealand	Roughly two-thirds Cabernet and one-third Merl...
129954	New Zealand	One of the more characterful Pinot Gris for th...
129955	New Zealand	Like Dog Point's 2011 Chardonnay, this wine is...
129956	New Zealand	The blend is 44% Merlot, 33% Cabernet Sauvigno...
129957	Spain	Lightly baked berry aromas vie for attention w...
129958	New Zealand	This blend of Cabernet Sauvignon-Merlot and Ca...
129959	France	The granite soil of the Brand Grand Cru vineya...
129960	Portugal	Fresh and fruity, this is full of red cherry f...
129961	Italy	Intense aromas of wild cherry, baking spice, t...
129962	Italy	Blackberry, cassis, grilled herb and toasted a...
129963	Israel	A bouquet of black cherry, tart cranberry and ...
129964	France	Initially quite muted, this wine slowly develo...
129965	France	While it's rich, this beautiful dry wine also ...
129966	Germany	Notes of honeysuckle and cantaloupe sweeten th...
129967	US	Citation is given as much as a decade of bottl...
129968	France	Well-drained gravel soil gives this wine its c...
129969	France	A dry style of Pinot Gris, this is crisp with ...
129970	France	Big, rich and off-dry, this is powered by inte...

		designation	points	price \
0		Vulkà Bianco	87	NaN
1		Avidagos	87	15.0
2		NaN	87	14.0
3		Reserve Late Harvest	87	13.0
4		Vintner's Reserve Wild Child Block	87	65.0
5		Ars In Vitro	87	15.0
6		Belsito	87	16.0
7		NaN	87	24.0
8		Shine	87	12.0
9		Les Natures	87	27.0
10		Mountain Cuvée	87	19.0
11		NaN	87	30.0
12		NaN	87	34.0
13		Rosso	87	NaN
14		NaN	87	12.0
15		Devon	87	24.0
16		Felix	87	30.0
17		Winemaker Selection	87	13.0
18	Vendimia Seleccionada Finca Valdelayegua Singl...		87	28.0
19		NaN	87	32.0
20		Vin de Maison	87	23.0
21		NaN	87	20.0
22		Ficiligno	87	19.0
23		Signature Selection	87	22.0
24		Aynat	87	35.0
25		King Ridge Vineyard	87	69.0
26		Dalila	87	13.0

27		NaN	87	10.0
28		Mascaria Barricato	87	17.0
29		NaN	86	16.0
...		...	...	...
129941		NaN	90	20.0
129942		NaN	90	35.0
129943		Adènzia	90	29.0
129944		Special Reserve Winemakers' Choice	90	25.0
129945		Jurassic Park Vineyard Old Vines	90	20.0
129946		Dom	90	17.0
129947		Symposio	90	20.0
129948		Pedernal	90	43.0
129949		Barrel Fermented	90	35.0
129950		Blocks 7 & 22	90	35.0
129951		Holder	90	30.0
129952		NaN	90	22.0
129953		Elspeth	90	25.0
129954		Single Estate	90	15.0
129955		NaN	90	40.0
129956	Gimblett Gravels Merlot-Cabernet Sauvignon-Malbec		90	19.0
129957		Crianza	90	17.0
129958		Irongate	90	35.0
129959		Brand Grand Cru	90	57.0
129960		Vértice	90	48.0
129961		NaN	90	30.0
129962		Sàgana Tenuta San Giacomo	90	40.0
129963		Oak Aged	90	20.0
129964		Domaine Saint-Rémy Herrenweg	90	NaN
129965		Seppi Landmann Vallée Noble	90	28.0
129966	Brauneberger Juffer-Sonnenuhr Spätlese		90	28.0
129967		NaN	90	75.0
129968		Kritt	90	30.0
129969		NaN	90	32.0
129970	Lieu-dit Harth Cuvée Caroline		90	21.0

	province	region_1	region_2 \
0	Sicily & Sardinia	Etna	NaN
1	Douro	NaN	NaN
2	Oregon	Willamette Valley	Willamette Valley
3	Michigan	Lake Michigan Shore	NaN
4	Oregon	Willamette Valley	Willamette Valley
5	Northern Spain	Navarra	NaN
6	Sicily & Sardinia	Vittoria	NaN
7	Alsace	Alsace	NaN
8	Rheinhessen	NaN	NaN
9	Alsace	Alsace	NaN
10	California	Napa Valley	Napa
11	Alsace	Alsace	NaN

12	California	Alexander Valley	Sonoma
13	Sicily & Sardinia	Etna	NaN
14	California	Central Coast	Central Coast
15	Mosel	NaN	NaN
16	Other	Cafayate	NaN
17	Mendoza Province	Mendoza	NaN
18	Northern Spain	Ribera del Duero	NaN
19	Virginia	Virginia	NaN
20	Virginia	Virginia	NaN
21	Oregon	Oregon	Oregon Other
22	Sicily & Sardinia	Sicilia	NaN
23	California	Paso Robles	Central Coast
24	Sicily & Sardinia	Sicilia	NaN
25	California	Sonoma Coast	Sonoma
26	Sicily & Sardinia	Terre Siciliane	NaN
27	Sicily & Sardinia	Terre Siciliane	NaN
28	Sicily & Sardinia	Cerasuolo di Vittoria	NaN
29	California	Clarksburg	Central Valley
...	...	...	...
129941	California	Mendocino County	NaN
129942	California	Sonoma County	Sonoma
129943	Sicily & Sardinia	Sicilia	NaN
129944	Galilee	NaN	NaN
129945	California	Santa Ynez Valley	Central Coast
129946	Mosel	NaN	NaN
129947	Sicily & Sardinia	Terre Siciliane	NaN
129948	Other	San Juan	NaN
129949	California	Napa Valley	Napa
129950	California	Napa Valley	Napa
129951	Alsace	Alsace	NaN
129952	California	Chiles Valley	Napa
129953	Hawke's Bay	NaN	NaN
129954	Marlborough	NaN	NaN
129955	Marlborough	NaN	NaN
129956	Hawke's Bay	NaN	NaN
129957	Northern Spain	Rioja	NaN
129958	Hawke's Bay	NaN	NaN
129959	Alsace	Alsace	NaN
129960	Douro	NaN	NaN
129961	Sicily & Sardinia	Sicilia	NaN
129962	Sicily & Sardinia	Sicilia	NaN
129963	Galilee	NaN	NaN
129964	Alsace	Alsace	NaN
129965	Alsace	Alsace	NaN
129966	Mosel	NaN	NaN
129967	Oregon	Oregon	Oregon Other
129968	Alsace	Alsace	NaN
129969	Alsace	Alsace	NaN



129970	Alsace	Alsace	NaN
--------	--------	--------	-----

	taster_name	taster_twitter_handle	\
0	Kerin OKeefe	@kerinokeefe	
1	Roger Voss	@vossroger	
2	Paul Gregutt	@paulgwineā	
3	Alexander Peartree	NaN	
4	Paul Gregutt	@paulgwineā	
5	Michael Schachner	@wineschach	
6	Kerin OKeefe	@kerinokeefe	
7	Roger Voss	@vossroger	
8	Anna Lee C. Iijima	NaN	
9	Roger Voss	@vossroger	
10	Virginie Boone	@vboone	
11	Roger Voss	@vossroger	
12	Virginie Boone	@vboone	
13	Kerin OKeefe	@kerinokeefe	
14	Matt Kettmann	@mattkettmann	
15	Anna Lee C. Iijima	NaN	
16	Michael Schachner	@wineschach	
17	Michael Schachner	@wineschach	
18	Michael Schachner	@wineschach	
19	Alexander Peartree	NaN	
20	Alexander Peartree	NaN	
21	Paul Gregutt	@paulgwineā	
22	Kerin OKeefe	@kerinokeefe	
23	Matt Kettmann	@mattkettmann	
24	Kerin OKeefe	@kerinokeefe	
25	Virginie Boone	@vboone	
26	Kerin OKeefe	@kerinokeefe	
27	Kerin OKeefe	@kerinokeefe	
28	Kerin OKeefe	@kerinokeefe	
29	Virginie Boone	@vboone	
...	...	...	
129941	Virginie Boone	@vboone	
129942	Virginie Boone	@vboone	
129943	Kerin OKeefe	@kerinokeefe	
129944	Mike DeSimone	@worldwineguys	
129945	Matt Kettmann	@mattkettmann	
129946	Anna Lee C. Iijima	NaN	
129947	Kerin OKeefe	@kerinokeefe	
129948	Michael Schachner	@wineschach	
129949	Virginie Boone	@vboone	
129950	Virginie Boone	@vboone	
129951	Roger Voss	@vossroger	
129952	Virginie Boone	@vboone	
129953	Joe Czerwinski	@JoeCz	
129954	Joe Czerwinski	@JoeCz	

129955	Joe Czerwinski	@JoeCz
129956	Joe Czerwinski	@JoeCz
129957	Michael Schachner	@wineschach
129958	Joe Czerwinski	@JoeCz
129959	Roger Voss	@vossroger
129960	Roger Voss	@vossroger
129961	Kerin OKeefe	@kerinokeefe
129962	Kerin OKeefe	@kerinokeefe
129963	Mike DeSimone	@worldwineguys
129964	Roger Voss	@vossroger
129965	Roger Voss	@vossroger
129966	Anna Lee C. Iijima	NaN
129967	Paul Gregutt	@paulgwineã
129968	Roger Voss	@vossroger
129969	Roger Voss	@vossroger
129970	Roger Voss	@vossroger

		title \
0	Nicosia 2013 Vulkà Bianco	(Etna)
1	Quinta dos Avidagos 2011 Avidagos Red	(Douro)
2	Rainstorm 2013 Pinot Gris	(Willamette Valley)
3	St. Julian 2013 Reserve Late Harvest Riesling ...	
4	Sweet Cheeks 2012 Vintner's Reserve Wild Child...	
5	Tandem 2011 Ars In Vitro Tempranillo-Merlot (N...	
6	Terre di Giurfo 2013 Belsito Frappato	(Vittoria)
7	Trimbach 2012 Gewurztraminer	(Alsace)
8	Heinz Eifel 2013 Shine Gewürztraminer	(Rheinhe...
9	Jean-Baptiste Adam 2012 Les Natures Pinot Gris...	
10	Kirkland Signature 2011 Mountain Cuvée Caberne...	
11	Leon Beyer 2012 Gewurztraminer	(Alsace)
12	Louis M. Martini 2012 Cabernet Sauvignon	(Alex...
13	Masseria Setteporte 2012 Rosso	(Etna)
14	Mirassou 2012 Chardonnay	(Central Coast)
15	Richard Böcking 2013 Devon Riesling	(Mosel)
16	Felix Lavaque 2010 Felix Malbec	(Cafayate)
17	Gaucha Andino 2011 Winemaker Selection Malbec ...	
18	Pradorey 2010 Vendimia Seleccionada Finca Vald...	
19	Quiévreumont 2012 Meritage	(Virginia)
20	Quiévreumont 2012 Vin de Maison Red	(Virginia)
21	Acrobat 2013 Pinot Noir	(Oregon)
22	Baglio di Pianetto 2007 Ficiligno White	(Sicilia)
23	Bianchi 2011 Signature Selection Merlot	(Paso ...
24	Canicattì 2009 Aynat Nero d'Avola	(Sicilia)
25	Castello di Amorosa 2011 King Ridge Vineyard P...	
26	Stemmari 2013 Dalila White	(Terre Siciliane)
27	Stemmari 2013 Nero d'Avola	(Terre Siciliane)
28	Terre di Giurfo 2011 Mascaria Barricato	(Cera...
29	Clarksburg Wine Company 2010 Chenin Blanc	(Cla...

... ..

129941           Apriori 2013 Chardonnay (Mendocino County)

129942   Arrowood 2010 Cabernet Sauvignon (Sonoma County)

129943   Baglio del Cristo di Campobello 2012 Adènzia R...

129944   Barkan 2011 Special Reserve Winemakers' Choice...

129945   Birichino 2013 Jurassic Park Vineyard Old Vine...

129946   Bischöfliche Weingüter Trier 2013 Dom Riesling...

129947   Feudo Principi di Butera 2012 Symposio Red (Te...

129948       Finca Las Moras 2010 Pedernal Malbec (San Juan)

129949   Flora Springs 2013 Barrel Fermented Chardonnay...

129950   Hendry 2012 Blocks 7 & 22 Zinfandel (Napa Valley)

129951   Henri Schoenheitz 2012 Holder Gewurztraminer (...)

129952           Houdini 2011 Zinfandel (Chiles Valley)

129953   Mills Reef 2011 Elspeth Cabernet Merlot (Hawke...

129954       Ara 2013 Single Estate Pinot Gris (Marlborough)

129955           Dog Point 2012 Chardonnay (Marlborough)

129956   Esk Valley 2011 Gimblett Gravels Merlot-Cabern...

129957           Viñedos Real Rubio 2010 Crianza (Rioja)

129958           Babich 2010 Irongate Red (Hawke's Bay)

129959   Cave de Turckheim 2010 Brand Grand Cru Pinot G...

129960   Caves Transmontanas 2006 Vértice Pinot Noir (D...

129961           COS 2013 Frappato (Sicilia)

129962   Cusumano 2012 Sàgana Tenuta San Giacomo Nero d...

129963   Dalton 2012 Oak Aged Cabernet Sauvignon (Galilee)

129964   Domaine Ehrhart 2013 Domaine Saint-Rémy Herren...

129965   Domaine Rieflé-Landmann 2013 Seppi Landmann Va...

129966   Dr. H. Thanisch (Erben Müller-Burggraef) 2013 ...

129967           Citation 2004 Pinot Noir (Oregon)

129968   Domaine Gresser 2013 Kritt Gewurztraminer (Als...

129969       Domaine Marcel Deiss 2012 Pinot Gris (Alsace)

129970   Domaine Schoffit 2012 Lieu-dit Harth Cuvée Car...

	variety	winery
0	White Blend	Nicosia
1	Portuguese Red	Quinta dos Avidagos
2	Pinot Gris	Rainstorm
3	Riesling	St. Julian
4	Pinot Noir	Sweet Cheeks
5	Tempranillo-Merlot	Tandem
6	Frappato	Terre di Giurfo
7	Gewürztraminer	Trimbach
8	Gewürztraminer	Heinz Eifel
9	Pinot Gris	Jean-Baptiste Adam
10	Cabernet Sauvignon	Kirkland Signature
11	Gewürztraminer	Leon Beyer
12	Cabernet Sauvignon	Louis M. Martini
13	Nerello Mascalese	Masseria Setteporte
14	Chardonnay	Mirassou

15	Riesling	Richard Böcking
16	Malbec	Felix Lavaque
17	Malbec	Gaucha Andino
18	Tempranillo Blend	Pradorey
19	Meritage	Quiévreumont
20	Red Blend	Quiévreumont
21	Pinot Noir	Acrobat
22	White Blend	Baglio di Pianetto
23	Merlot	Bianchi
24	Nero d'Avola	Canicatti
25	Pinot Noir	Castello di Amorosa
26	White Blend	Stemmari
27	Nero d'Avola	Stemmari
28	Red Blend	Terre di Giurfo
29	Chenin Blanc	Clarksburg Wine Company
...	...	...
129941	Chardonnay	Apriori
129942	Cabernet Sauvignon	Arrowood
129943	Red Blend	Baglio del Cristo di Campobello
129944	Shiraz	Barkan
129945	Chenin Blanc	Birichino
129946	Riesling	Bischöfliche Weingüter Trier
129947	Red Blend	Feudo Principi di Butera
129948	Malbec	Finca Las Moras
129949	Chardonnay	Flora Springs
129950	Zinfandel	Hendry
129951	Gewürztraminer	Henri Schoenheit
129952	Zinfandel	Houdini
129953	Cabernet Merlot	Mills Reef
129954	Pinot Gris	Ara
129955	Chardonnay	Dog Point
129956	Bordeaux-style Red Blend	Esk Valley
129957	Tempranillo Blend	Viñedos Real Rubio
129958	Bordeaux-style Red Blend	Babich
129959	Pinot Gris	Cave de Turckheim
129960	Pinot Noir	Caves Transmontanas
129961	Frappato	COS
129962	Nero d'Avola	Cusumano
129963	Cabernet Sauvignon	Dalton
129964	Gewürztraminer	Domaine Ehrhart
129965	Pinot Gris	Domaine Rieflé-Landmann
129966	Riesling	Dr. H. Thanisch (Erben Müller-Burggraef)
129967	Pinot Noir	Citation
129968	Gewürztraminer	Domaine Gresser
129969	Pinot Gris	Domaine Marcel Deiss
129970	Gewürztraminer	Domaine Schoffit

[129971 rows x 13 columns]

```
In [122]: reviews.groupby("country").country.count();#replica of value counts
```

```
In [123]: #How many times each point occurs appears  
reviews.groupby("points").points.count()
```

```
Out[123]: points  
80      397  
81      692  
82     1836  
83     3025  
84     6480  
85     9530  
86    12600  
87    16933  
88    17207  
89    12226  
90    15410  
91    11359  
92     9613  
93     6489  
94     3758  
95     1535  
96      523  
97      229  
98       77  
99       33  
100      19  
Name: points, dtype: int64
```

```
In [127]: #What is the minimum price for each point...i.e  
#Cheapest wine in each point value category  
reviews.groupby("points").price.min()
```

```
Out[127]: points  
80      5.0  
81      5.0  
82      4.0  
83      4.0  
84      4.0  
85      4.0  
86      4.0  
87      5.0  
88      6.0  
89      7.0  
90      8.0  
91      7.0  
92     11.0  
93     12.0  
94     13.0
```

```

95      20.0
96      20.0
97      35.0
98      50.0
99      44.0
100     80.0
Name: price, dtype: float64

```

```
In [128]: reviews.groupby("winery").winery.count()
```

```

Out[128]: winery
1+1=3                6
10 Knots             4
100 Percent Wine     3
1000 Stories         2
1070 Green           1
10Span              4
12 Linajes           8
12C Wines            5
13 Celsius           1
14 Hands            67
16X20               2
1752 Signature Wines 1
1789 Wines           1
181                 1
18401 Cellars        5
1848 Winery          6
1850                3
19 Crimes            1
1912 Winemakers      8
2 Copas              1
2 Lads               7
2 Up                4
21 Grams             2
24 Knots             1
240 Days             3
25 Lagunas           1
29 & Oak Wines       5
2Hawk                7
2Plank               6
2nd Chance           3
..
histoire             1
iNapa Wine           1
j.brix               3
kukkula              10
l'Escargot           1
m2 Wines             1

```

majoros	2
next:	1
noblet	1
oops	13
pb	1
the Ghost of 413	2
un4seen	5
una.wine	3
Ànima Negra	14
Áldás	1
Árido	5
Áster	1
Çamliba	2
Ébano	3
Écluse	21
Élevée Winegrowers	3
Éric & Jöel Durand	4
Ïl Macchione	2
Ñandú	1
Órale	1
Öko	2
Ökonomierat Rebholz	4
àMaurice	40
toka	3

Name: winery, Length: 16757, dtype: int64

In [130]: *#What does this do?*  
 reviews.groupby("winery").apply(lambda df:df.title.iloc[0])

Out[130]: winery

1+1=3	1+1=3 NV Rosé Sparkling (Cava)
10 Knots	10 Knots 2010 Viognier (Paso Robles)
100 Percent Wine	100 Percent Wine 2015 Moscato (California)
1000 Stories	1000 Stories 2013 Bourbon Barrel Aged Zinfande...
1070 Green	1070 Green 2011 Sauvignon Blanc (Rutherford)
10Span	10Span 2013 Pinot Noir (Central Coast)
12 Linajes	12 Linajes 2009 Reserva (Ribera del Duero)
12C Wines	12C Wines 2012 Beckstoffer Vineyard Georges I...
13 Celsius	13 Celsius 2014 Sauvignon Blanc (Marlborough)
14 Hands	14 Hands 2010 Riesling (Washington)
16X20	16X20 2010 Chardonnay (Sonoma Coast)
1752 Signature Wines	1752 Signature Wines 2016 Marion Rosé (Côtes d...
1789 Wines	1789 Wines 2007 Pinot Noir (Chehalem Mountains)
181	181 2008 Merlot (Lodi)
18401 Cellars	18401 Cellars 2013 Cabernet Sauvignon (Walla W...
1848 Winery	1848 Winery 2009 Reserve Cabernet Merlot (Gali...
1850	1850 2015 Rosé (Amador County)
19 Crimes	19 Crimes 2013 Red (South Eastern Australia)

1912 Winemakers	1912 Winemakers 2011 Vinhas Velhas Branco Whit...
2 Copas	2 Copas 2009 Red (Mendoza)
2 Lads	2 Lads 2009 Cabernet Franc-Merlot (Old Mission...
2 Up	2 Up 2006 Shiraz (South Australia)
21 Grams	21 Grams 2011 Red (Washington)
24 Knots	24 Knots 2009 Pinot Noir (California)
240 Days	240 Days 2015 Riesling (Finger Lakes)
25 Lagunas	25 Lagunas 2014 Cabernet Sauvignon (San Juan)
29 & Oak Wines	29 & Oak Wines 2014 Cabernet Sauvignon (Napa V...
2Hawk	2Hawk 2014 Cabernet Franc (Rogue Valley)
2Plank	2Plank 2014 Zinfandel (Amador County)
2nd Chance	2nd Chance 2009 Pinot Noir (Santa Maria Valley)
	...
histoire	histoire 2006 Chapter 1 Cabernet Sauvignon (Me...
iNapa Wine	iNapa Wine 2009 Lillian Vineyard Estate Grown ...
j.brix	j.brix 2013 The Augur Kick On Ranch Vineyard R...
kukkula	kukkula 2014 Noir Syrah (Adelaida District)
l'Escargot	l'Escargot 2013 Sauvignon Blanc (Côtes de Gasc...
m2 Wines	m2 Wines 2014 Pinot Noir (Los Carneros)
majoros	majoros 2013 Furmint (Tokaj)
next:	next: 2008 Riesling (Washington)
noblet	noblet 2014 Côtes de Bourg
oops	oops 2011 Red (Central Valley)
pb	pb 2010 Heritage Pinot Gris (Yakima Valley)
the Ghost of 413	the Ghost of 413 2010 Red (Columbia Valley (WA))
un4seen	un4seen 2011 Chardonnay (Lodi)
una.wine	una.wine 2015 Late Harvest Chardonnay (Burgenl...
Ànima Negra	Ànima Negra 2004 Àn Red (Vi de la Terra Illes ...
Áldás	Áldás 2005 Egri Bikavér Red (Eger)
Árido	Árido 2013 Sustainably Farmed Moscato (Mendoza)
Áster	Áster 2010 Finca El Otero (Ribera del Duero)
Çamliba	Çamliba 2012 Karalahna (Turkey)
Ébano	Ébano 2007 6 (Ribera del Duero)
Écluse	Écluse 2014 Rendition Red (Paso Robles)
Élevée Winegrowers	Élevée Winegrowers 2012 Élevée Vineyard Pinot ...
Éric & Jöel Durand	Éric & Jöel Durand 1999 Cornas
Īl Macchione	Īl Macchione 2008 Riserva (Vino Nobile di Mon...
Ñandú	Ñandú 2010 Malbec (Mendoza)
Órale	Órale 2011 Cabronita Red (Santa Ynez Valley)
Öko	Öko 2013 Made With Organically Grown Grapes Ma...
Ökonomierat Rebholz	Ökonomierat Rebholz 2007 Von Rotliegenden Spät...
àMaurice	àMaurice 2013 Fred Estate Syrah (Walla Walla V...
toka	toka 2009 Izbrani Teran (Kras)

Length: 16757, dtype: object

```
In [151]: #You can group by more than one column
reviews.groupby(["country", "province"]).count()
```

```
Out[151]:
```

	description	designation	points	price	\
--	-------------	-------------	--------	-------	---



country	province				
Argentina	Mendoza Province	3264	2500	3264	3226
	Other	536	402	536	530
Armenia	Armenia	2	2	2	2
Australia	Australia Other	245	144	245	236
	New South Wales	85	50	85	85
	South Australia	1349	1040	1349	1331
	Tasmania	42	27	42	42
	Victoria	322	195	322	315
	Western Australia	286	184	286	285
Austria	Austria	26	20	26	22
	Burgenland	672	461	672	567
	Carnuntum	103	80	103	96
	Eisenberg	16	12	16	16
	Kamptal	409	384	409	338
	Kremstal	340	326	340	292
	Leithaberg	42	6	42	34
	Mittelburgenland	23	19	23	19
	Neusiedlersee	13	7	13	11
	Niederösterreich	600	516	600	481
	Steiermark	20	19	20	14
	Südburgenland	2	2	2	2
	Südoststeiermark	37	35	37	27
	Südsteiermark	109	99	109	78
	Thermenregion	128	95	128	107
	Traisental	43	42	43	43
	Vienna	72	61	72	58
	Wachau	353	350	353	308
	Wagram	65	59	65	62
	Wagram-Donauland	56	54	56	34
	Weinland Österreich	18	15	18	16
...		...	...	...	...
US	Hawaii	1	1	1	1
	Idaho	192	132	192	190
	Illinois	6	6	6	6
	Iowa	4	3	4	4
	Kentucky	1	1	1	1
	Massachusetts	7	4	7	7
	Michigan	114	89	114	111
	Missouri	33	23	33	33
	Nevada	4	1	4	4
	New Jersey	8	7	8	8
	New Mexico	45	27	45	45
	New York	2688	1867	2688	2676
	North Carolina	23	14	23	23
	Ohio	12	10	12	12
	Oregon	5373	3789	5373	5359
	Pennsylvania	18	11	18	18

	Rhode Island	1	1	1	1
	Texas	94	77	94	93
	Vermont	3	3	3	3
	Virginia	777	457	777	770
	Washington	8639	5892	8639	8583
	Washington-Oregon	7	5	7	7
Ukraine	Ukraine	14	12	14	14
Uruguay	Atlantida	5	5	5	5
	Canelones	43	37	43	43
	Juanico	12	8	12	12
	Montevideo	11	8	11	11
	Progreso	11	11	11	11
	San Jose	3	3	3	3
	Uruguay	24	23	24	24

		region_1	region_2	taster_name \
country	province			
Argentina	Mendoza Province	3211	0	3261
	Other	536	0	536
Armenia	Armenia	0	0	2
Australia	Australia Other	245	0	196
	New South Wales	85	0	76
	South Australia	1347	0	1183
	Tasmania	42	0	33
	Victoria	322	0	302
	Western Australia	286	0	247
Austria	Austria	0	0	18
	Burgenland	0	0	672
	Carnuntum	0	0	103
	Eisenberg	0	0	16
	Kamptal	0	0	409
	Kremstal	0	0	340
	Leithaberg	0	0	42
	Mittelburgenland	0	0	23
	Neusiedlersee	0	0	13
	Niederösterreich	0	0	600
	Steiermark	0	0	20
	Südburgenland	0	0	2
	Südoststeiermark	0	0	37
	Südsteiermark	0	0	109
	Thermenregion	0	0	128
	Traisental	0	0	43
	Vienna	0	0	72
	Wachau	0	0	353
	Wagram	0	0	65
	Wagram-Donauland	0	0	56
	Weinland Österreich	0	0	18
...		...	...	...

US	Hawaii	1	0	1
	Idaho	28	0	187
	Illinois	6	0	6
	Iowa	4	0	3
	Kentucky	1	0	1
	Massachusetts	7	0	3
	Michigan	114	0	106
	Missouri	33	0	4
	Nevada	4	0	3
	New Jersey	8	0	5
	New Mexico	45	0	24
	New York	2688	2688	2597
	North Carolina	23	0	15
	Ohio	12	0	8
	Oregon	5372	5067	5337
	Pennsylvania	18	0	13
	Rhode Island	1	0	0
	Texas	85	0	85
	Vermont	3	0	3
	Virginia	777	0	379
	Washington	8639	8637	8631
	Washington-Oregon	0	0	7
Ukraine	Ukraine	0	0	14
Uruguay	Atlantida	0	0	5
	Canelones	0	0	43
	Juanico	0	0	12
	Montevideo	0	0	11
	Progreso	0	0	11
	San Jose	0	0	3
	Uruguay	0	0	24

		taster_twitter_handle	title	variety	winery
country	province				
Argentina	Mendoza Province	3261	3264	3264	3264
	Other	536	536	536	536
Armenia	Armenia	2	2	2	2
Australia	Australia Other	196	245	245	245
	New South Wales	76	85	85	85
	South Australia	1183	1349	1349	1349
	Tasmania	33	42	42	42
	Victoria	302	322	322	322
	Western Australia	247	286	286	286
Austria	Austria	18	26	26	26
	Burgenland	672	672	672	672
	Carnuntum	103	103	103	103
	Eisenberg	16	16	16	16
	Kamptal	409	409	409	409
	Kremstal	340	340	340	340

	Leithaberg	42	42	42	42
	Mittelburgenland	23	23	23	23
	Neusiedlersee	13	13	13	13
	Niederösterreich	600	600	600	600
	Steiermark	20	20	20	20
	Südburgenland	2	2	2	2
	Südoststeiermark	37	37	37	37
	Südsteiermark	109	109	109	109
	Thermenregion	128	128	128	128
	Traisental	43	43	43	43
	Vienna	72	72	72	72
	Wachau	353	353	353	353
	Wagram	65	65	65	65
	Wagram-Donauland	56	56	56	56
	Weinland Österreich	18	18	18	18
...	...	...	...	...	...
US	Hawaii	0	1	1	1
	Idaho	186	192	192	192
	Illinois	0	6	6	6
	Iowa	2	4	4	4
	Kentucky	1	1	1	1
	Massachusetts	1	7	7	7
	Michigan	2	114	114	114
	Missouri	2	33	33	33
	Nevada	0	4	4	4
	New Jersey	3	8	8	8
	New Mexico	15	45	45	45
	New York	243	2688	2688	2688
	North Carolina	1	23	23	23
	Ohio	6	12	12	12
	Oregon	5337	5373	5373	5373
	Pennsylvania	12	18	18	18
	Rhode Island	0	1	1	1
	Texas	8	94	94	94
	Vermont	3	3	3	3
	Virginia	17	777	777	777
	Washington	8631	8639	8639	8639
	Washington-Oregon	7	7	7	7
Ukraine	Ukraine	12	14	14	14
Uruguay	Atlantida	5	5	5	5
	Canelones	43	43	43	43
	Juanico	12	12	12	12
	Montevideo	11	11	11	11
	Progreso	11	11	11	11
	San Jose	3	3	3	3
	Uruguay	24	24	24	24

[425 rows x 11 columns]

```
In [156]: #What is the best wine by country and province? Wine with the highest points
reviews.groupby(["country", "province"]).apply(lambda df: df.loc[df.points.idxmax]);
```

Try it yourself: Find the average price of wine per country

```
In [198]: reviews.groupby("country").price.mean()
```

```
Out[198]: country
Argentina      24.510117
Armenia         14.500000
Australia      35.437663
Austria        30.762772
Bosnia and Herzegovina 12.500000
Brazil         23.765957
Bulgaria       14.645390
Canada         35.712598
Chile          20.786458
China          18.000000
Croatia        25.450704
Cyprus         16.272727
Czech Republic 24.250000
Egypt          NaN
England        51.681159
France         41.139120
Georgia        19.321429
Germany        42.257547
Greece         22.364425
Hungary        40.648276
India          13.333333
Israel         31.768916
Italy          39.663770
Lebanon        30.685714
Luxembourg     23.333333
Macedonia      15.583333
Mexico         26.785714
Moldova        16.745763
Morocco        19.500000
New Zealand    26.931785
Peru           18.062500
Portugal       26.218256
Romania        15.241667
Serbia         24.500000
Slovakia       16.000000
Slovenia       24.812500
South Africa   24.668987
Spain          28.215275
Switzerland    85.285714
Turkey         24.633333
```

US	36.573464
Ukraine	9.214286
Uruguay	26.403670

Name: price, dtype: float64

Exercise: 1. How many reviews did each reviewer do? 2. What is the best wine i can buy for a given amount of money? 3. What are the minimum and maximum prices for each variety of wine? 4. What is the average rating given by each reviewer? 5. How many times does each combination of country and variety appear?

```
In [165]: #1
          reviews.groupby("taster_twitter_handle").taster_twitter_handle.count();

In [169]: #2
          reviews.groupby("price").points.max();

In [173]: #3
          reviews.groupby("variety").price.min();

In [175]: #3
          reviews.groupby("variety").price.max();

In [177]: reviews.groupby("variety").price.agg([min,max]);

In [181]: #4
          values=reviews.groupby("taster_name").points.mean();

In [202]: #5
          reviews.groupby(["country","variety"]).size(); #Could also use count
```

### 1.3.7 Data Types and Missing Data

Data type for a **column** in a DataFrame or Series is known as dtype. You can use the dtype attribute to determine the type of a specific column. dtypes returns the dtype of every column in the data set You can convert a column type with the astype function.

```
In [203]: reviews.price.dtype
```

```
Out[203]: dtype('float64')
```

```
In [205]: reviews.dtypes
```

```
Out[205]: country          object
          description      object
          designation      object
          points           int64
          price            float64
          province         object
          region_1         object
          region_2         object
```

```

taster_name          object
taster_twitter_handle object
title                object
variety              object
winery               object
dtype: object

```

```
In [207]: reviews.points.astype("float64");
```

**Missing Data** Missing values are shown as NaN, short for Not a Number. NaN values are always of the float64 dtype Pandas provides methods specific to missing data. To select NaN entries you can use `pd.isnull()` or its companion `pd.notnull()`

```
In [211]: #reviews where the country is NaN
reviews[reviews.country.isnull()];
```

```
In [212]: #reviews where the country is not NaN
reviews[reviews.country.notnull()];
```

Replacing missing values is a common operation. Pandas provides `fillna()` method for dealing with missing values

```
In [218]: #replace each missing region 2 with unknown
reviews.region_2.fillna("Unknown"); #There are quite a number of imputing strategies
```

Non-null values can be replaced. E.g suppose that since this dataset was published reviewer Kering O'Keefe has changed her Twitter handle from `@kerinokeefe` to `@kerino`. You can use the `replace` method to reflect this in the dataset. This can be an important method in replacing missing data given some kind of sentinel value in the data e.g. unknown, invalid etc

```
In [229]: reviews.taster_twitter_handle.replace("@kerinokeefe", "@kerino");
```