**Lesson 11**

Grab data from multiple excel files and merge them into a single dataframe.

In [1]:

**import** **pandas** **as** **pd**

**import** **os**

**import** **sys**

%**matplotlib** inline

In [2]:

print 'Python version ' + sys.version

print 'Pandas version: ' + pd.\_\_version\_\_

Python version 2.7.5 |Anaconda 2.1.0 (64-bit)| (default, Jul 1 2013, 12:37:52) [MSC v.1500 64 bit (AMD64)]

Pandas version: 0.15.2

**Create 3 excel files¶**

In [3]:

*# Create DataFrame*

d = {'Channel':[1], 'Number':[255]}

df = pd.DataFrame(d)

df

Out[3]:

|  |  |  |
| --- | --- | --- |
|  | **Channel** | **Number** |
| **0** | 1 | 255 |

In [4]:

*# Export to Excel*

df.to\_excel('test1.xlsx', sheet\_name = 'test1', index = **False**)

df.to\_excel('test2.xlsx', sheet\_name = 'test2', index = **False**)

df.to\_excel('test3.xlsx', sheet\_name = 'test3', index = **False**)

print 'Done'

Done

**Place all three Excel files into a DataFrame¶**

Get a list of file names but make sure there are no other excel files present in the folder.

In [5]:

*# List to hold file names*

FileNames = []

*# Your path will be different, please modify the path below.*

os.chdir(r"C:\Users\david\notebooks\pandas")

*# Find any file that ends with ".xlsx"*

**for** files **in** os.listdir("."):

**if** files.endswith(".xlsx"):

FileNames.append(files)

FileNames

Out[5]:

['test1.xlsx', 'test2.xlsx', 'test3.xlsx']

Create a function to process all of the excel files.

In [6]:

**def** GetFile(fnombre):

*# Path to excel file*

*# Your path will be different, please modify the path below.*

location = r'C:\Users\david\notebooks\pandas**\\**' + fnombre

*# Parse the excel file*

*# 0 = first sheet*

df = pd.read\_excel(location, 0)

*# Tag record to file name*

df['File'] = fnombre

*# Make the "File" column the index of the df*

**return** df.set\_index(['File'])

Go through each file name, create a dataframe, and add it to a list.

i.e. df\_list = [df, df, df]

In [7]:

*# Create a list of dataframes*

df\_list = [GetFile(fname) **for** fname **in** FileNames]

df\_list

Out[7]:

[ Channel Number

File

test1.xlsx 1 255, Channel Number

File

test2.xlsx 1 255, Channel Number

File

test3.xlsx 1 255]

In [8]:

*# Combine all of the dataframes into one*

big\_df = pd.concat(df\_list)

big\_df

Out[8]:

|  |  |  |
| --- | --- | --- |
|  | **Channel** | **Number** |
| **File** |  |  |
| **test1.xlsx** | 1 | 255 |
| **test2.xlsx** | 1 | 255 |
| **test3.xlsx** | 1 | 255 |

In [9]:

big\_df.dtypes

Out[9]:

Channel int64

Number int64

dtype: object

In [10]:

*# Plot it!*

big\_df['Channel'].plot(kind='bar');

