

Instructions

For today's activity you will work with your team to complete the Pandas tutorial at: <http://synesthesiam.com/posts/an-introduction-to-pandas.html>

A few notes about this tutorial. It is a good tutorial that highlights some important elements of the pandas library, but it is using a slightly older version than we are. You will likely notice this in the output of some commands being similar, but slightly different, which is just fine and should not alarm you. Also, you'll see the use of some things like: `data.irow(3)` which [has been deprecated](#) (i.e., the old way phased out) in favor of `data.iloc[3]`.

Also, depending on how your environment is configured, running a plotting command may not display the plot automatically. There are various things you can do such as saving this to a file, etc., but you might consider importing `matplotlib.pyplot` directly and then calling `.show()` on it. For example:

```
import matplotlib.pyplot as plt
```

```
... other things here ...
```

```
data.mean_temp.hist()
```

```
# if nothing pops up, you can then do:
```

```
plt.show()
```

As with any programming project, but especially when using libraries, the internet is your friend. If something doesn't quite work the way you expect, you should turn to an internet search to see what others are saying and doing.

CORE REQUIREMENTS

1. Use the pip or pip3 installer to install pandas, numpy, and matplotlib. (Numpy will come automatically with Pandas)

Instructor Tip:

If you have trouble installing it via pip on the command line, you could instead use PyCharm to download and install this package.

To do so, with your project open in PyCharm, get to the project settings page. (On a Mac this is under the PyCharm -> Preferences menu option.) Expand the Project:xxxx item in the list and select "Project Interpreter" (where xxxx is the name of your project). Click the "+" button at the bottom of the page that lists the packages, then type "pandas" in the search box, select it, and click "Install". This should install pandas for use in your project.

2. Download the weather dataset from the tutorial webpage, open an interactive Python console window. Ensure that you can load the dataset with pandas as follows:

```
import pandas
data = pandas.read_csv("weather_year.csv")
data
```

Instructor Tip:

If you type the above commands directly into the interpreter, you'll see output from the last one ("data"). However, if you run a script with the above code, it will not display anything, because none of those lines produce output. To make the last line print out a summary of the dataset to the console, you'll need to instead have "print(data)".

This is because when you run commands directly in the interpreter, it will display the result of the expression automatically, but running a program won't. Hence the difference between just saying "x" to see the value of the variable x, verses having to say "print(x)".

3. Go through each step of the tutorial. Make sure each member of the team stays together on each step.

To complete this requirement, you should complete everything up through and including the section "Fun with Columns". This includes getting the histogram to display.

Stretch Challenges

After completing the above steps. Make sure that everyone on the entire team is to this point and understands the material. Then, if you have time, move on to the following stretch challenges.

1. Complete the sections:
 - Bulk operations with `apply()`
 - Handling Missing Values
 - Accessing Individual Rows
2. Complete the sections:
 - Filtering
 - Grouping
 - Creating New Columns
3. Complete the sections:
 - Plotting
 - Getting Data Out
 - Miscellanea