Pandas cookbook

pandas is a Python library for doing data analysis. It's really fast and lets you do exploratory work incredibly quickly.

The goal of this cookbook is to give you some concrete examples for getting started with pandas. The docs are really comprehensive. However, I've often had people tell me that they have some trouble getting started, so these are examples with real-world data, and all the bugs and weirdness that entails.

I'm working with 3 datasets right now

- 311 calls in New York
- How many people were on Montréal's bike paths in 2012
- Montreal's weather for 2012, hourly

It comes with batteries (data) included, so you can try out all the examples right away.

Table of Contents

- Chapter 5: Combining dataframes and scraping Canadian weather data
  Here you get to find out if it's cold in Montreal in the winter (spoiler: yes). Web scraping with pandas is fun!
- Chapter 6: String operations! Which month was the snowiest?
  Strings with pandas are great. It has all these vectorized string operations and they're the best. We will turn a bunch of strings containing "Snow" into vectors of numbers in a trice.
- Chapter 7: Cleaning up messy data
  Cleaning up messy data is never a joy, but with pandas it's easier <3
- Chapter 8: Parsing Unix timestamps
  This is basically a quick trick that took me 2 days to figure out.
- Chapter 9 - Loading data from SQL databases
  How to load data from an SQL database into Pandas, with examples using SQLite3, PostgreSQL, and MySQL.

How to use this cookbook

The easiest way is to try it out instantly online using Binder's awesome service. **Start by clicking here**, wait for it to launch, then click on "cookbook", and you'll be off to the races! It will let you run all the code interactively without having to install anything on your computer.

To install it locally , you'll need an up-to-date version of IPython Notebook (>= 3.0) and n your computer pandas (>=0.13) for this to work properly. It's set up to work with Python 2.7.

You can get these using `pip` (you may want to do this inside a virtual environment to avoid conflicting with your other libraries).

```
  pip install -r requirements.txt
```

This can be difficult to get set up and require you to compile a whole bunch of things. I instead use and recommend Anaconda, which is a Python distribution which will give you everything you need. It's free and open source.

Once you have pandas and IPython, you can get going!

```
git clone https://github.com/jvns/pandas-cookbook.git
cd pandas-cookbook/cookbook
ipython notebook
```

A tab should open up in your browser at `http://localhost:8888`
Happy pandas!

Running the cookbook inside Docker container.

This repository contains Dockerfile and can be built into a docker container. To build the container run following command from inside of the repository directory:

```
docker build -t jvns/pandas-cookbook -f Dockerfile-Local .
```

run the container:

```
docker run -d -p 8888:8888 -e "PASSWORD=MakeAPassword" <IMAGE ID>
```

you can find out about the id of the image, by checking

```
docker images
```

After starting the container, you can access ipython notebook with the cookbook on port 8888. Remember to use https and authenticate with `MakeAPassword`.

```
https://<docker ip>:8888
```

Contribute!

If you see something wrong, or there's something you'd like to learn that I haven't explained here, or there's something you know about that you would like to share, create an issue! Send me email! Send a pull request!