Resources

Read or watch:

- 10 minutes to Pandas
- Complete Python Pandas Data Science Tutorial (Reading CSV/Excel files, Sorting, Filtering, Groupby)

Learning Objectives

At the end of this project, you are expected to be able to explain to anyone, without the help of Google:

- What is pandas?
- What is a pd.DataFrame? How do you create one?
- What is a pd.Series? How do you create one?
- How to load data from a file
- How to perform indexing on a pd.DataFrame
- How to use hierarchical indexing with a pd.DataFrame
- How to slice a pd.DataFrame
- How to reassign columns
- How to sort a pd.DataFrame
- How to use boolean logic with a pd.DataFrame
- How to merge/concatenate/join pd.DataFrames
- How to get statistical information from a pd.DataFrame
- How to visualize a pd.DataFrame

Requirements

General

- Allowed editors: `vi`, `vim`, `emacs`
- All your files will be interpreted/compiled on Ubuntu 16.04 LTS using `python3` (version 3.5)
- Your files should be executed with `numpy` (version 1.15) and `pandas` (version 0.24)
- All your files should end with a new line
- The first line of all your files should be exactly `#!/usr/bin/env python3`
- All of your files must be executable
- A `README.md` file, at the root of the folder of the project, is mandatory
- Unless otherwise noted, you can only use `import pandas as pd`
- You are scheduled to follow the pep8 style (version 2.4)
- All your modules should have documentation (`python3 -c 'print(__import__("my_module").__doc__)'`)
- All your classes should have documentation (`python3 -c 'print(__import__("my_module").MyClass.__doc__)'`)
- All your functions (inside and outside a class) should have documentation (`python3 -c 'print(__import__("my_module").my_function.__doc__)'` and `python3 -c 'print(__import__("my_module").MyClass.my_function.__doc__)'`)

Download Pandas 0.24.x

- pip install --user pandas

Datasets

For this project, we will be using the `coinbase` and `bitstamp` datasets, as seen previously in 0x0E: Time Series Forecasting.