

PyData NYC 2017 Summary

(Two days tutorial only)

Each presenter had materials posted on their respective github account. None of them were able to go through their entire presentation given their allotted time. This event is hosted by Microsoft at their Time Square technology Center.

Develop Interactive Matplotlib figures

Thomas A Caswell

Matplotlib lead developer and Physicist at BNL

Attended both sessions (3h).

Thomas prepared 7 examples to demonstrate how to interact with your plot based on button pressed, key pressed and other events. Being able to create or filter data interactively is highly desirable for exploratory application. Some python script have hands on exercises to try out different events.

Python script are available on his github account:

https://github.com/tacaswell/interactive_mpl_tutorial

Installation notes are in README file. Demonstration was done using IPython console in a dedicated environment. Once you activated your new environment, you can start IPython and run each file like this:

```
%run -i SCRIPTNAME.py
```

Same code can be run in Jupyter notebook by adding this line:

```
%matplotlib notebook
```

Each python script can be run in Spyder if you install spyder in your newly created environment.

Network Science, Game of Thrones and US Airports

Mridul Seth

Graduate Student, developer for NetworkX

Brief introduction to networkx module. Examples showing most popular Game of Thrones characters in each book. Also studied commercial US domestic flights to find out the average number of connections needed between all airports.

Jupyter notebooks are available at this link:

<https://github.com/MridulS/pydatanyc-2017-networkx>

Instructions are in the README file

Pandas: .head() to .tail()

Tom Augspurger

Data Scientist, Anaconda

Attended session 2 out of 2 (1h30).

Basic introduction on selecting/indexing data, tidy data as well as grouped, merge and join operations. It also shows some data preprocessing on beer reviews data and it does it lazily.

Jupyter notebooks are available here:

<https://github.com/tomaugspurger/pydata-nyc-ph2t>

Installation notes are in the README file

Idiomatic Pandas

Ted Petrou

Founder of Dunder Data (Training services)

Attended sessions 1 and 2 out of 2 (3h total)

Very experienced and energetic presenter! Showed how to use pandas efficiently. Use native pandas function instead of native python function. Avoid using apply unless there is no work around to vectorize your operation. Use .loc and .iloc for creating a subset data. He also describe when pandas create a view versus a copy.

Jupyter notebook are available here:

<https://github.com/tdpetrou/Learn-Pandas>

No particular version of Pandas necessary.

Bayesian Statistics from Scratch: Building up to MCMC

Justin Bozonier

Data Scientist at GrubHub

MCMC stands for Markov chain Monte Carlo. He started from a very basic bayes theorem and finished by showing a MCMC example.

Just a single Jupyter notebook at :

<https://github.com/jcbozonier/research/tree/master/notebooks/PyDataNYC2017.ipynb>

Have the following installed on your computer:

Python 3

IPython & Jupyter Notebooks

Numpy

Pymc3

Scipy
Matplotlib
pandas

Loan Default Modeling in Python: An Introductory Guide

Lore Dirick

Recent PhD graduate working at Data Camp

She did her PhD in Belgium on credit default prediction. She ran through a dataset and showed data preparation before applying logistic regression and its associated ROC. She didn't have time to show her decision tree model (spoiler: logistic model did better). Instead there were more interest in the survival analysis.

All Jupyter notebook are available at:

https://github.com/LoreDirick/PyData_2017

Required modules are in the README, one should create a new environment if you have any missing modules.

After the last tutorial session, I had a chance to talk to Ted Petrou (presenter on pandas). He is a former poker player who turned Data Scientist after starting a pandas meetup group in Austin, TX. He is now offering training courses on pandas.

Ted also introduced me to Paige Bailey, cloud developer advocate in Machine Learning/AI for Microsoft. I connected with both on LinkedIn.