Python for Data Analysis

YETI

Agenda

1. Welcome, introductions & objectives
2. What is Python and when should I use it?
3. Tutorials
   - Python Basics
   - Pandas Basics
4. Visualizations
5. Next steps & reference materials

Introduction

Course Objectives

1. Understand what Python is and when to use it
2. Get Python installed and running on your computer
3. **Get hands on experience using Python**
4. Receive a list of resources to continue learning

Agenda

1. Welcome, introductions & objectives
2. What is Python and when should I use it?
3. Tutorials
   - Python Basics
   - Pandas Basics
4. Visualizations
5. Next steps & reference materials

What is Python?

Multi-purpose programming language	**Easy to learn** (replaced Java at colleges)	**Readable** (indents instead of brackets, dynamic typing)	**Open-sourced**
Versatile ("Python is generally the second-best language for everything.")	**Large community** (many tutorials available and questions answered on stack overflow)	**Alternatives**: R, SAS, Scala	**Not** SQL

When to use Python

AUTOMATING REPETITIVE TASKS

- Combining PDFs
- Searching for text in files
- Downloading information from the internet
- Merging/updating spreadsheets
- Sending emails
- Aggregating multiple data sources

DATA ANALYSIS

- Creating reproducible analyses and reports
- Making visualizations and graphics
- Interacting with databases and using SQL
- Statistical analysis
- Transforming data (e.g. pivot/filter/summarize)

MORE ADVANCED FEATURES

- Creating web pages
- Machine learning / AI
- Manipulating images / sound

Excel or Python?

EXCEL

- Good for **small, quick** analyses
- Allows you to **more easily see the data**
- Presenting **tables of data** and **non-complex charts**

PYTHON

Use when Excel struggles:

- Lots of **data**
- Lots of **formulas**, in particular:
  - lookups
  - matches
  - sumifs
  - array formulas
- Functions that either **don't exist** in Excel or are **difficult to use**:
  - Unpivoting/melting
  - Histograms
- If you are using **VBA**
- Performing **repeatable processes** (avoid manual errors!!)

Tutorials

Choose as many tutorials that interest you.

AVAILABLE TUTORIALS

- Stock Prices
- Manipulate Fortune Data
- Visualizations
- Basic Webscraping
- Database connections
- Working with PDFs
- Create Excel Files

TUTORIALS IN DEVELOPMENT

- Combining Multiple Files
- Sending Emails

Note: You can find each tutorial in the course files previously downloaded.

Agenda

1. Welcome, introductions & objectives
2. What is Python and when should I use it?
3. Tutorials
    - Python Basics
    - Pandas Basics
4. Visualizations
5. Next steps & reference materials

Course Files

You should have received an email with a zip folder. To access course files…

1. Save the zip file to your laptop and unzip it.
2. Copy “Course Files” folder and paste into your Downloads folder
   - The path to the Workbooks folder should now look like this: C:\Users\yourusername\Downloads\Course Files\Workbooks
3. Open command prompt: windows key -> cmd -> enter
4. Type “jupyter notebook” and hit enter

Python Basics

Now, we'll look at the basics of Python.

In Jupyter Notebook open 'AFU PDA 1 – Python Basics.ipynb'

Pandas Basics

Now, we'll look at basic Pandas functionality.

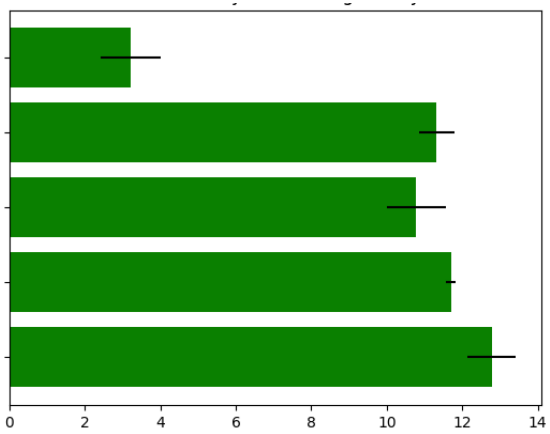
In Jupyter Notebook open 'AFU PDA 2 - Pandas basics.ipynb'

Agenda

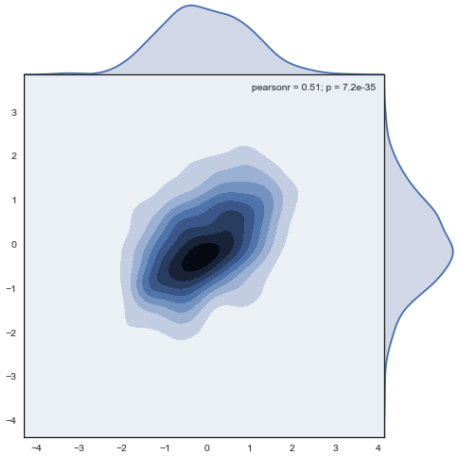
1. Welcome, introductions & objectives
2. What is Python and when should I use it?
3. Tutorials
   - Python Basics
   - Pandas Basics
4. Visualizations
5. Next steps & reference materials

Visualizations

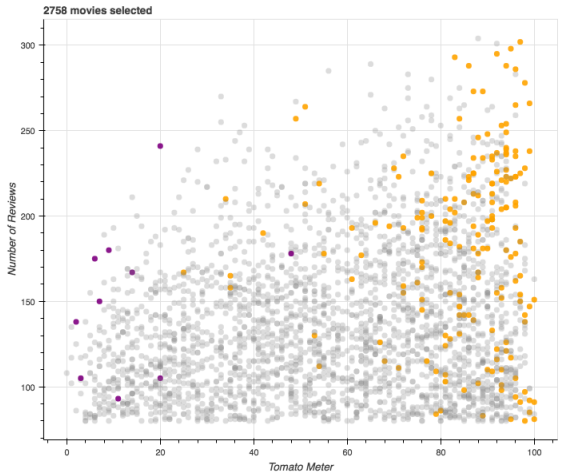
Matplotlib



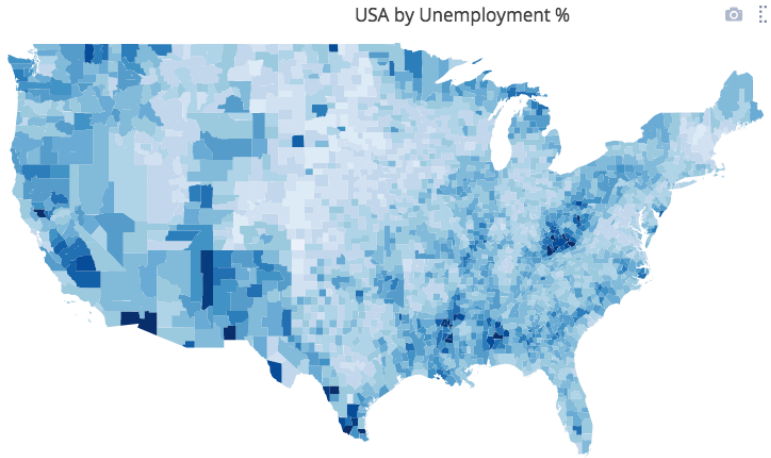
Seaborn



Bokeh



Plotly



Learn more at
Afu.amazon.com

Agenda

1. Welcome, introductions & objectives
2. What is Python and when should I use it?
3. Tutorials
   - Python Basics
   - Pandas Basics
4. Visualizations
5. Next steps & reference materials

Next Steps and Reference Materials

1. Mailing List: AFU-python-for-data-analysis
2. Office Hours: Sign up here
3. Stack Overflow – forum for programming troubleshooting
4. Python Data Science Handbook – Series of notebooks covering many data science topics
5. Udacity: Intro to Data Analysis
6. Lynda: Python: Data Analysis
7. Automate the Boring Stuff with Python
8. Codeacademy