

Data Analysis Using Python: A Beginner's Guide Featuring NYC Open Data

Mark Bauer

March 6, 2021

NYC Open Data Week 2021

The aim of this presentation is to walk you through my project located on my GitHub.

Opinions expressed are solely my own and do not express the views or opinions of my employer.



This is not a workshop. We will not be live coding.



Prerequisites

- Basics of Python or other programming languages (R, SQL, etc.)
- Knowledge of Data Analysis
- Basics of Jupyter Notebooks

But those new to Python are gladly welcome!



Accessibility

- Volume and speed (e.g. am I speaking too loudly?)
- Mute your microphone
 - Does everyone know how to mute and unmute on Zoom?
- Participation
 - Use the chat feature and/or raise your hand feature
 - If speaking during question breaks, please announce yourself



Access the Slides on my GitHub

<https://github.com/mebauer/data-analysis-using-python>



Background Project Advice to Beginners Preview on GitHub



Background

Project

Advice to Beginners

Preview on GitHub



I created this beginner series because



One Can Learn Python

- University
- On the Job
- Self-Study



One Can Learn Topics

- Statements
- Data Types
- Functions and Methods
- Loops

```
print("Hello World!")
```



But there exists a disconnect between learning and applying Python for
your own projects



My Journey

- Numerous online courses (Coursera, EdX, Udemy)
- Attended tech meetups
- Books

However, I was still unclear how to apply skills to actual projects



NYU and Professor Bianco

- Pursued in an Advanced Certificate Program at NYU in Urban Science and Informatics
- Enrolled in Professor Federica Bianco's *Principles of Urban Informatics* (PUI) Class (Fall 2018)



Principles of Urban Informatics (PUI) Class

- Hands-on exercises
- Typical Data Analysis workflows
- Reproducible and open-source teaching
- Python focused on the library pandas

Changed my life



Background

Project

Advice to Beginners

Preview on GitHub



Project on GitHub

mebauer / data-analysis-using-python

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 0 tags Go to file Add file Code

Mark Bauer adding links for open source applications used in project 2007d8f 3 days ago 177 commits

1-reading-writing-files adding links to static webpages for notebooks 3 days ago

2-data-inspection-cleaning-wrang... adding links to static webpages for notebooks 3 days ago

3-plotting-data-visualizations adding links to static webpages for notebooks 3 days ago

4-geospatial-data-mapping adding links to static webpages for notebooks 3 days ago

mybinder-examples adding hexbin map to examples in binder 6 days ago

.gitignore add .gitignore file 4 months ago

CODE_OF_CONDUCT.md Create CODE_OF_CONDUCT.md 3 months ago

CONTRIBUTING.md revising contributing.md for typos 3 months ago

LICENSE Create LICENSE 13 months ago

README.md adding links for open source applications used in project 3 days ago

environment.yml editing yaml file 7 days ago

README.md

Data Analysis Using Python: A Beginner's Guide Featuring NYC Open Data

Mark Bauer

About

Data Analysis Using Python: A Beginner's Guide Featuring NYC Open Data

python nyc numpy
pandas-dataframe
exploratory-data-analysis
jupyter-notebook pandas seaborn
open-data data-analytics
data-analysis matplotlib
python-tutorial socrata geopandas
pandas-tutorial socrata-open-data-api
nyc-opendata nyc-open-data

Readme

MIT License

Releases

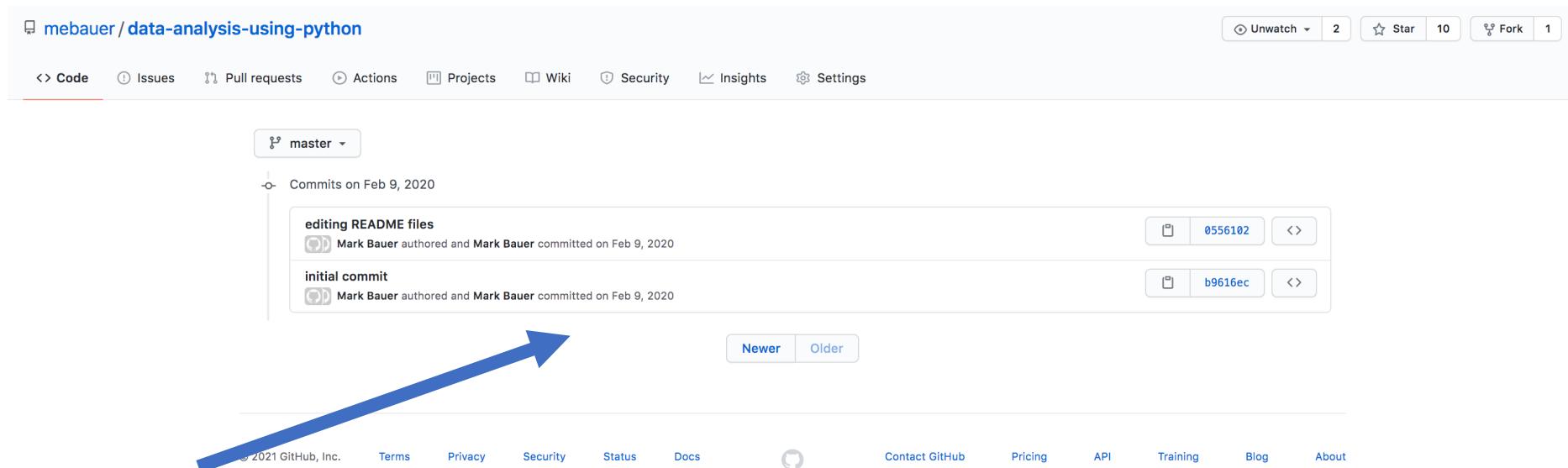
No releases published Create a new release

Packages

No packages published Publish your first package



A Long Journey



First commit on 2/9/20



You Will Learn How To

- Read (load) and write (export) data
- Inspect, process, and wrangle data
- Create data visualizations
- Work with geospatial data and create maps



You Will Learn (Indirectly)

- Develop a presence on GitHub to showcase work
- Create beautiful notebooks
- Learn markdown language and create informative Readme.md files
- Think about your audience
- ***Reproducibility*** and ***open-source research***



You Will Need

- Tools
 - Python
 - Jupyter Notebook (accessed through Anaconda)
- Knowledge
 - Beginner-level proficiency with Python (and familiarity with packages such as pandas, numpy, matplotlib, seaborn)
 - Basic knowledge of Data Analysis (e.g. descriptive statistics)



Background Project Advice to Beginners Preview on GitHub



Before we go to the project on GitHub



Advice for Beginners

- Install Anaconda and write code in a Jupyter Notebook
- Learn the Python library pandas
- Review examples of other people's code
- Learn to enjoy coding (debugging is not always fun 😕)
- Begin to produce your own projects

A bit later: learn Git (version control) and GitHub



Background Project Advice to Beginners Preview on GitHub



Preview the project on GitHub

Data Analysis Using Python: A Beginner's Guide Featuring NYC Open Data

<https://github.com/mebauer/data-analysis-using-python>

- Additional Links:
 - Run an interactive example through your browser (takes ~5 minutes to load): [mybinder](#)
 - View *Data Inspection and Wrangling* notebook as static webpage: [nbviewer](#)

