

Introduction to Python for Data Analysis

Professor Jason Anastasopoulos

Personal Webpage: <http://anastasopoulos.io>

Email: ljasonanastas@gmail.com

Course Preliminaries

Before the first day of class please:

- Create a free Google Drive account: <https://www.google.com/drive/>
- Familiarize yourself with Google Colaboratory Python Notebooks:
<https://colab.research.google.com/notebooks/intro.ipynb>
- Join our Slack Channel:
https://join.slack.com/t/intro-to-python-ch/shared_invite/zt-xcdkiept-ccTZWsLYGM5omSNmyZEMxw
- Download and Install Anaconda Python 3.7+: <https://www.anaconda.com/products/distribution>

Google Drive Materials*

https://drive.google.com/drive/folders/1ojXGTIKKHfAtnMAWDtuxRT-pNrxap_5?usp=sharing

*Check each day before class for updates to code and slides.

Tentative Course Schedule¹

Day 1: Introduction to Python

06/21/2022

Time	Activity	Topic	Content
Before June 21st	Optional Prep ▾	<i>Recommended Reading</i>	<u>Python for Everybody</u> by Dr. Charles R. Severance • Chapters 1-6, 8, 9. • Link: https://www.py4e.com/html3/
Make an editable copy for yourself. Review before live session.	Colab Notebook ▾	<i>Day 1 Notebook: Introduction to Python</i>	• Day 1 Notebook
10:30-12:30PM (EST)	Live session ▾	Getting started with Python	• Why Python? • Introduction to Anaconda Python. • Introduction to Python Google/Jupyter notebooks. • Overview of basic libraries used: NumPy, Pandas, Matplotlib, SciPy, statsmodels. • Variables: numbers, strings values, using variables.
12:30-1:30	Break ▾		
1:30-3:00PM	Live session ▾	Python basics & data structures	• Lists and loops: lists basics, simple loops, pythonic loops. • Logical statements in python. • Using and creating dictionaries. • Creating functions.

¹ Exact content covered for each time slot will vary.

Day 2: Data Wrangling

06/22/2022

Time	Activity	Topic	Content
After Day 1, Before Day 2	Optional Prep ▾	Recommended Reading	Python Data Science Handbook (Free Google Colab Notebooks) by Jake VanderPlas Data Import/Export with Pandas • Introduction to Pandas • Introducing Pandas Objects • Data Indexing and Selection
Make an editable copy for yourself. Review before live session.	Colab Notebook ▾	Day 2 Notebook 2: Data Wrangling	• Day 2 Notebook.
10:30-12:30PM (EST)	Live session ▾	Introduction to NumPy	• Arrays: manipulation and transformation. • Important functions.
	Live session ▾	Introduction to Pandas	• Import/Export data with Pandas • Data Indexing with Pandas
12:30-1:30	Break ▾		
1:30-3:00PM	Live session ▾	Advanced Data Manipulation with Pandas	• Advanced data frame creation and manipulation. • Sorting data, replacing data. • Dealing with missing values. • Data queries.
	Live session ▾	Q & A Session I*	Question and Answer Session

*Time permitting

Day 3: Data Analysis

06/23/2022

Time	Activity	Topic	Content
Make an editable copy for yourself. Review before live session.	Colab Notebook ▾	Day 3 Notebook: Data Analysis	• Day 3 Notebook
10:30-12:30PM (EST)	Live session ▾	Data Analysis	• Summary statistics: mean, median, mode, variance and standard deviation. • Hypothesis testing: t-tests, confidence intervals. • Linear and logistic regression. • Regression diagnostics.
12:30-1:30	Break ▾		
1:30-3:00PM	Live session ▾	Q&A Session II*	Question and answer session.

* Time permitting.

Day 4: Data Visualization

06/23/2022

Time	Activity	Topic	Content
Make an editable copy for yourself. Review before live session.	Colab Notebook ▾	Day 4 Notebook: <i>Data Visualization</i>	• Day 4 Notebook
10:30-12:30PM (EST)	Live session ▾	Intro to Visualization and Advanced Visualization	• Intro to data visualization in Python using <i>matplotlib</i> and <i>seaborn</i>. • Distributions: densities, box plots, histograms. • Correlations: scatterplots, line plots, heatmaps.
12:30-1:30	Break ▾		
1:30-3:00PM	Live session ▾	Q&A Session III*	Bring any questions that you have about the past four days

* Time permitting.