Data Analysis Using Python

Instructor

Brandon Krakowsky, Lecturer at the School of Engineering and Research & Education Director at the Wharton School

Course Description

This course provides an introduction to basic data science techniques using Python. Students are introduced to core concepts like Data Frames and joining data, and learn how to use data analysis libraries like pandas, *numpy*, and *matplotlib*. This course provides an overview of loading, inspecting, and querying real-world data, and how to answer basic questions about that data. Students will gain skills in data aggregation and summarization, as well as basic data visualization.

Course Learning Objectives

- Apply basic data science techniques using Python
- Understand and apply core concepts like Data Frames and joining data, and use data analysis libraries like pandas, numpy, and matplotlib
- Demonstrate how to load, inspect, and query real-world data, and answer basic questions about that data
- Analyze data further by applying learned skills in data aggregation and summarization, as well as basic data visualization

Intended Audience

This course is intended for students and professionals who have minimal or no prior programming exposure. It's for motivated learners who have experience with rigorous coursework, and are looking to gain a competitive edge in advancing their career.

Course Prerequisites

- High school or college math.
- "Introduction to Python Programming" Coursera course or equivalent prior knowledge of introductory Python.

Course Outline

Module 1: Loading, Querying, & Filtering Data Using the csv Module

- Learning Objectives
  - Review Jupyter Notebook and basic Python
  - Apply basic data science techniques using Python
  - Demonstrate how to load, read, and explore data with Python's csv module
  - Select and sort data based on criteria
- Topics
  - Loading, Querying, & Filtering .csv Files
  - Catching Data Errors & Sorting

Module 2: Loading, Querying, Joining & Filtering Data Using pandas

- Learning Objectives
  - Understand and apply core concepts like Data Frames and joining data, and use the pandas data analysis library
  - Demonstrate how to load, inspect, and query real-world data, and answer basic questions about that data
  - Select specific data using advanced queries and filters
  - Create new data
- Topics
  - Loading, Inspecting, & Querying Data
  - Joining & Filtering Data

Module 3: Summarizing & Visualizing Data

- Learning Objectives
  - Use data analysis libraries like numpy and matplotlib
  - Analyze data by applying skills in data aggregation and summarization
  - Group data and calculate meaningful summary statistics
  - Visualize data in Jupyter Notebook
- Topics
  - Summarizing Data
  - Visualizing Data

Course Assessment

This course will use a variety of assessments. Ungraded code-along videos allow students to practice along with the instructor, and self-assess their ability to apply the concepts and skills they learned, before attempting the graded assessments. Graded assessments include:

- Quizzes to check your knowledge in each module (except in Module 1)

- Programming assignments to test your level of understanding

To earn a certificate in this course, learners must earn a passing score on all assessments:

- Homework Assignments: 60% or above
- Quizzes: 75% or above

Recommended Resources

- *Python Crash Course*, by Eric Matthes: https://nostarch.com/pythoncrashcourse2e
  (You should also be able to find it on the O'Reilly site through the UPenn library)
- *Think Python*, by Allen B. Downey: https://greenteapress.com/wp/think-python/
- *Automate the Boring Stuff with Python*, by Al Sweigart: https://automatetheboringstuff.com/
- *Python in Easy Steps*, by Mike McGrath: https://www.amazon.com/Python-easy-steps-Covers-3-7/dp/1840788127/

Effort

We expect this course will take you 5-7 hrs per week to complete, for a total of 3 weeks.

Communication and Support

You can communicate with course staff and other students through the discussion forums. Please reach out to us through the discussion forum with any questions about the course content. Please allow at least 48 hours to receive a response from a TA or course staff.

Note: All communication on the discussion forum must follow the Coursera Honor Code . Never post code or solutions to assignments on the discussion forum. If you are having difficulty with code or solutions, a TA may provide an email address to send it in for private assistance.