CMU ONLINE
Carnegie Mellon University

Syllabus Overview Sample: Python for Data Science I

Program: Machine Learning & Data Science Certificate Program

Faculty: Dr. Ralf Brown, Principal Systems Scientist, Language Technologies Institute

Course Number: 11904 **Units:** 6

Zoom Class: Weekday Evening, ET

Course Description

This is the first mini of a two-course sequence on the Python programming language, with a focus on learning concepts, techniques, skills, and tools needed for developing programs in Python. In this course, students gain hands-on experience solving real-world problems by completing programming projects in Python focused on the language fundamentals, software development practices, data manipulation and analysis. Specifically, students are exposed to realistic software development projects, real-world data, and scenarios.

Course Learning Objectives

By the end of this course, students will be able to:

- Explain and use fundamental constituents of the Python programming language.
- Use the Python language to solve a variety of computational problems.
- Analyze and visualize real-world data using Python data science libraries.

Faculty Introduction



Dr. Ralf Brown is a Principal Systems Scientist at Carnegie Mellon University's Language Technologies Institute (LTI). With a Ph.D. in Computer Science from Carnegie Mellon, Dr. Brown has been a pivotal member of the institute since 1993, contributing to groundbreaking advancements in machine translation, language identification, and computational linguistics. His distinguished career has been recognized with awards such as the IJCAI Distinguished Paper Award and Carnegie Mellon's Allen Newell Award for Research Excellence.

Dr. Brown's research spans the development of innovative tools and techniques in example-based machine translation, corpus linguistics, and low-resource language processing. He is the creator of the CMU-EBMT, an open-source example-based machine translation system, and the LTI LangID Corpus, a resource for training language identifiers in over 2,000 languages. His work combines theoretical insights with practical applications, advancing the frontiers of natural language processing in diverse and multilingual contexts.

A dedicated educator and mentor, Dr. Brown leads courses on algorithm design and computational linguistics, emphasizing hands-on learning and real-world problem-solving. He has repeatedly served as Chair of the PhD Admissions Committee. He also actively contributes to open-source projects, including the widely used darktable image editor. With decades of experience and a passion for linguistic innovation, Dr. Brown continues to inspire the next generation of researchers at LTI while driving progress in language technologies worldwide.

Course Topics

Topic	Learning Objectives
Intro to Python and Jupyter	● Execute code in a jupyter cell, and complete an assignment using our Canvas Jupyter environment. ● Create Python Statements and Expressions by combining variables and operators.
Stings, Functions and Flow Control	● Design and Evaluate Flow Control in Python. ● Construct functions for particular tasks. ● Read data from a file using with and open.
Strings and Data Structures	● Implement data structures in python, including lists, strings, sets, and dictionaries. ● Choose the appropriate data structure, and distinguish between immutable and mutable types. ● Manipulate strings using python's built-in string operations.
JSON and Regular Expressions	● Search and Modify strings using regular expressions. ● Create, Load and Modify data in JSON format
Object Oriented Programming	● Explain what object-oriented programming is, describe its core principles, and contrast the approach to procedural programming. ● Explain and implement examples of prominent special methods in Python, such as init, str, len, and eq. ● Distinguish between private and public methods and variables. ● Modify and use a python class to process data (phone numbers).
Files and Formatting	● Use the OS and shutil modules to interact with the file system and operating system. ● Interact with the system from a Jupyter notebook.

	● Use exception handling and flow control to deal with unexpected or missing inputs to a function or script. ● Create a shell script as a standalone file (not a notebook) that take command line inputs (sys.argv)
Pandas	● Create and export plots using matplotlib. ● Use and contrast both matplotlib interfaces, the object and the plot interface. ● Modify plots to look better and better communicate information.

NOTE: This is a sample designed to provide an overview of topics covered in the course based on previous course syllabi. All students will be provided an updated and final syllabus at the start of the course. Faculty, content, and schedule are all subject to change.