

Course Title: Introduction to Python

Course Number: G440

Credits: 1

Course Directors: Teon Brooks, PhD brookst3@mskcc.org

Course Prerequisites: None; Open to first year Cancer Biology and Engineering PhD students

Grading Policy: Pass/Fail

Course Description and Learning Objectives

This curriculum has been structured to emphasize the foundations in Python and its four fundamental scientific computing libraries: numpy , scipy , pandas , and matplotlib . This course also emphasizes two premier packages for statistical analyses and machine learning. This course wraps up with a capstone project to bring together all of these concepts in a practical and applied way.

Course Structure

The course meets for 10 sessions from 2:00 pm – 3:30 pm for six weeks. There will also be a final project session from 1:30-5:00. Changes to that schedule are communicated to students via email. Active learning and classroom pair programming will be equally used and integrated to complement traditional lectures.

Teaching Fellows

Teaching Fellows, drawn from senior GSK students and the postdoctoral community at MSK, are present in the course sessions. Their role is to act as an additional source of information/assistance, to help keep the discussion sessions moving.

Assignments and Methods for Assessing Student Achievement

This course is graded on a pass/fail basis; class participation and a final project will form the basis of the grading.

Course Evaluation

Students are expected to complete surveys regarding the lectures and overall course via their student portal. This feedback will be used to evaluate the effectiveness and relevance of the topics and provide direction for the subsequent iterations of the course.

Academic Dishonesty, Plagiarism and Artificial Intelligence

The Policy can be found in the [Student and Faculty Handbook](#) linked on the GSK Website.

Course Schedule

The course schedule can be found on Moodle.

Getting Started

Before class, please follow the instruction for installing the conda-forge distribution of Python ([conda-forge | community-driven packaging for conda](#)) We will spend the first class getting you all set with environments and additional tooling. If you have any problems with this installation, feel free to email at the address above.

Week 1

Module 1: Intro to Python

September 29, 2025

H1

1. Class overview
 1. Introductions
 2. Review syllabus
2. Setting up Python environment
 1. Installation
 2. What is the command line
 3. Text Editor vs. Jupyter Notebooks vs. IDE (e.g. Positron, Spyder)
2. What is Python
 1. Scripting Programming Language
 2. Background and Context
2. Familiarity with the environment and some basic Hello World things

H2

1. Interactive shell
2. Python data types
 1. Str, int, float, list, dict
3. Built-in functions and libraries

Module 2: Control Flow, Functions, and Classes

October 1, 2025

H1

1. Control Flow (If/Else, For loops)
2. Functions

H2

1. Objects/Classes
2. Bringing it all together

Week 2

Module 3: Intro to Numpy and Scipy

October 6, 2025

H1

1. Intro to Numpy
 1. NDarray
 2. Numeric operations

H2

1. Intro to Scipy - Basic stats
 1. Correlation
 2. T-tests

Module 4: Intro to Matplotlib and Seaborn

October 8, 2025

H1

1. Building a plot in Matplotlib
2. Building subplots in Matplotlib
3. Intro to Seaborn

H2

1. In-class exercise
 2. Homework Assignment
-

Week 3

Module 5: Intro to Pandas

October 15, 2025

H1

1. Dataframes

2. Loading Data
3. Summary Statistics

H2

1. Grouping and Aggregating

Module 6: Pandas, continued

October 17, 2025

Week 4

Module 7: Data Science in Practice

October 20, 2025

H1

1. Using Pandas for Statistics
2. Statsmodels
3. Re-visit Scipy.stats

H2

1. Data analysis report

Module 8: Intro to Sklearn

October 22, 2025

Week 5

Module 9: Advanced Topic

October 27, 2025

Module 10: In-class work on Final Project

October 29, 2025

Module 11: Final Project Presentation

November 5, 2025