

Course Title: Data Science with Python

Course Number: BU MET CS 677

Course Format: On-Line

Instructor Name: Eugene Pinsky
epinsky@bu.edu

Computer Science Department,
Metropolitan College, Boston University
1010 Commonwealth Avenue, Room 327
Boston, MA 02215

Course Times: Tue: 8:00 – 9:30 pm EST
Thu: 8:00 – 9:30 am EST

Office hours: by appointment

Course Description

There is a growing need for specialists with a background in Python who can apply data science methods to practical problems at their workplace. Working in data science requires an understanding of many interdisciplinary concepts, and involves data mining and application of various methods.

The proposed course is designed to fill this need. Students will learn major Python tools and techniques for data science and machine learning. There are weekly assignments and projects on topics covered in class. These assignments will help build necessary statistical, visualization, and other data science skills for effective use of data science in various applications, including finance, text processing, time series analysis, and

recommendation systems. In addition, students will choose a topic in data science for a final project.

Students can take the proposed course with not exclusively computer science backgrounds who have basic knowledge of Python.

Books

Required:

“Python for Data Analysis”, by W. McKinney, O’Reilly Publishing, 2017 (2nd edition), ISBN-13: 978-1491957660, purchased from Barnes & Noble

Recommended:

“Python Data Analysis” by Armando Fandango, Packt Publishing, ISBN-13: 978-1787127487

“Python Data Science Handbook” by Jake VanderPlas, O’Reilly Publishing, ISBN-13: 978-1491912058

Courseware: Blackboard, Course Notes

Class Policies

1. Assignment Completion & Late Work

Weekly programming assignments and quizzes submitted through blackboard on-line. Late homework is accepted with 50% penalty. Final projects are submitted through blackboard on-line. Both quizzes and final are closed-book.

2. Academic Conduct Code

Cheating and plagiarism will not be tolerated in any Metropolitan College course. They will result in no credit for the assignment or examination and may lead to disciplinary actions. Please take the time to review the Student Academic Conduct Code:

Academic conduct code as specified below:

http://www.bu.edu/met/metropolitan_college_people/student/resources/conduct/code.html.

Grading Criteria

35% homework, 20% quizzes, 30% final, 15% final project

Class Meetings, Lectures & Assignments

Week	Topic	Readings Due	Assignments
1	Review of Python, Numpy and data analysis libraries	Chapters 1,2 Course notes	Homework 1 Quiz 1
2	Pandas, Matplotlib & Seaborn, error metrics, model selection trade-offs	Chapter 4, 5, 8 Course notes	Homework 2 Quiz 2
3	Supervised learning and decision boundaries. Logistic regression and nearest neighbor classifiers. Parameter Estimation with gradient descent	Course notes	Homework 3 Quiz 3
4	Linear and polynomial models for prediction. Linear regression and classification. Parameter estimation	Course notes	Homework 4 Quiz 4



5	Bayes rule and Naïve Bayesian Classification. Decision trees. Ensemble learning with random forest classifiers	Course notes	Homework 4 Quiz 4
6	Large-margin classification and kernels. Support Vector Machines. Unsupervised learning. k -means clustering	Course notes	Homework 5 Quiz 5
7	Course review, project presentations and final exam	Course notes	Homework 6 Quiz 6