Boston University Metropolitan College

Course Title: Data Science with Python

Course Number: BU MET CS 677
Course Format: On Campus/Blended

Instructor Name: Eugene Pinsky
epinsky@bu.edu

Computer Science Department, Metropolitan College, Boston University

1010 Commonwealth Avenue, Room 327, Boston, MA 02215

Course Times: Monday 6:00 – 8:45, MET 101

Teaching Assistants:
1. Alisha Peermohamed (alisha22@bu.edu)
2. Gagan Kaushal (gk@bu.edu)

Office hours: Tuesdays 3-5 or by appointment

Course Description

At the present time, there is a growing need for specialists with 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, involves data mining and application of various methods.

The proposed course is designed to fill this need. Students will learn major Python tools, machine learning classifiers and techniques for data analysis. There are weekly assignments and mini 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 a variety of applications including finance, time series analysis and recommendation

systems. In addition, students will choose a topic for a final project and present it on the last day of class.

The proposed course can be taken by students 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 (2-nd 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

Additional materials will be added to "From Your Professor" section under group discussion section.

Class Policies

Weekly programming assignments submitted through blackboard on-line. Late homework is accepted with 50% penalty. Final projects are submitted through blackboard on-line. Students will present their projects on the last day of class. Both quiz and final are closed-book and in-class

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.
>
> NOTE: [This should not be understood as a discouragement for discussing the material or your particular approach to a problem with other students in the class. On the contrary – you should share your thoughts, questions and solutions

Grading Criteria:

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

Class Meetings, Lectures & Assignments:

The course is divided into 6 modules (each module is 2 weeks).

Module	Topic	Readings Due
1	Review of Python, Numpy and data analysis libraries	Chapters 1,2 Course notes
2	Pandas, Matplotlib & Seaborn, error metrics, model selection trade-offs	Chapter 4, 5, 8 Course notes

3	Supervised learning and decision boundaries. Logistic regression and nearest neighbor classifiers. Parameter Estimation with gradient descent	Course notes
4	Linear and polynomial models for prediction. Linear regression and classification. Parameter estimation	Course notes
5	Bayes rule and Naïve Bayesian Classification. Decision trees. Ensemble learning with random forest classifiers	Course notes
6	Large-margin classification and kernels. Support Vector Machines. Unsupervised learning. k-means clustering	Course notes
7	Course review, project presentations and final exam	Course notes