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

Time/Place: Wed 6 pm – 8:45 pm
COM Room 213

Office hours: Thursday 4-5:30 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 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, text processing, 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.

Boston University Metropolitan College

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 (including Review of Python notes)

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 are 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 as specified below:

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

Grading Criteria:

30% homework, 20% quizzes, 30% final, 15% final project. 5% discussion

Class Meetings, Lectures & Assignments:

Date	Topic	Readings Due
Week 1	Review of Python and data analysis libraries	Chapters 1,2 Course notes
Week 2	IPython Computing and Development Environment	Chapter 3 Course notes
Week 3	Numeric Python (Numpy) and vectorized computations, arrays, matrices, indexing, searching and sorting	Chapter 4 Course notes
Week 4	Pandas functionality, structures and operations, indexing and selection, data aggregation and grouping, pivot tables	Chapter 5 Course notes
Week 5	statistics and linear algebra with Python	Course notes
Week 6	data retrieving, outlier detection, cleaning and reprocessing (files, JSON, HTML)	Chapter 6, 7 Course notes
Week 7	plotting and visualization (matplotlib, line, scatter, barchart plots, histograms, subplots)	Chapter 8
Week 8	Data aggregation and group operations, working with data bases (SQL and no-SQL data retrieval with Pandas and SQLAlchemy	Chapter 9 Course notes
Week 9	signal processing and time series analysis, shifting and windowing (linear regression, moving averages, filtering, outlier detection, autocorrelations)	Chapter 10
Week 10	financial applications (financial time-series, price/volatility forecasting, trading algorithms)	Chapter 11
Week 11	analyzing textual and social media (analyzing frequencies, bag-of-words modeling, sentiment analysis)	Course notes
Week 12	predictive analytics (forecasting, classification with logistic regression, decision trees, clustering)	Course notes
Week 13	collaborative filtering and recommender systems (user and item-based filtering, similarity measures, nearest-neighbor methods)	Course notes
Week 14	Final Exam Project Presentations	