Analytics in Python

ColumbiaX BAMM.101x Micromaster

About this course

With the increase in both access to data as well as the tools available to deal with data, data analytics has become an exciting career choice and the purpose of this course is to prepare you with the basic tools necessary for data analysis. We will quickly learn the basics of the programming language Python, and then survey the tools available in Python for getting, cleaning, and analyzing data. We will see how to get data from files (csv, html, json, xml) and databases (Mysql), cover the rudiments of data cleaning, and examine data analysis, machine learning and data visualization packages (numpy, pandas, Scikit-learn) available in Python. We will, very briefly, look at the natural language processing and network analysis available in Python.

Prerequisites

We will review basic Python programming concepts in week 1 and 2 and no prior programming experience is necessary. But,if you have had some exposure to programming you will get more out of this class. There are many python programming resources online (for example, EdX has a good basic python class).

No specific math capability is required though it will be helpful if you are familiar with the basic concepts of algebra, set theory, and probability.

Week-by-week

Week 1: Python review
Week 2: Python review
Week 3: Data interchange formats: JSON and XML
Week 4: Web scraping and web crawling
Week 5: Database basics: Relational databases
Week 6: SQL
Week 7: Data analysis and visualization I
Week 8: Data analysis and visualization II
Week 9: Text mining
Week 10: Analysis of networks
Week 11: Machine learning: Part 1
Week 12: Machine learning: Part 2

Last update: 2017/06

Analytics in Python

ColumbiaX BAMM.101x Micromaster

GRADING

Quizzes (25%): The course has 16 quizzes.

Assignments (55%): The course has 7 assignments.

Final Exam (20%): There will be a final exam one week after the last lecture.

PASSING GRADE

To pass the course, you must score **60% or above**.

HONOR CODE

Academic Honesty Policy

You are required to read, and understand the following agreement regarding **Academic Honesty**. Each student is soleowner of his own code and work and **must NOT**:

- Submit work that is not original.
- Publish code or solutions online.
- Post the course questions on forums including stack overflow.
- Submit someone else's work, or a modification of that work, with or without that person's knowledge.
- Allow someone else to submit his/her work, or a modification of that work.
- Solve as a group a quiz or project. **All coursework is to be done by the student working alone.**
- Contract course work out to others.
- Plan or execute with another student a cooperative subterfuge during an exam.
- Make use of unauthorized material during an exam.

Project assignments will be checked with plagiarism detection software. Thank you for abiding by these rules. Doing so will ensure the experience is fair to everyone taking this class or the future sessions of this class.

Last update: 2017/06