Location	SEC 218
Date and Time	Wednesday 6:40PM - 9:30PM
Instructor	Ji Zheng ji.zheng1@rutgers.edu
Office Hours	5:30-6:30PM Wednesday
Text	*Python for Eve*ryone, Second Edition, by Cay S. Horstmann and Rance D. Necaise, Wiley, 2016 *Learning Python*, 5th Edition, by Mark Lutz, O'Reilly Media, 2013, ISBN-13: 978-1449355739.

Course Objectives

The Python programming language is an increasingly popular tool for the manipulation and analysis of data. course aims to give students the fundamental Python knowledge necessary for more advanced work in data science. The course assumes students already familiar with fundamental python program language syntax and program design. This course will build upon and reinforce these concepts and give students insight into how a large piece of software is built, and give students experience in managing a full-cycle development project. Part of the course is devoted to some popular Python packages for data analysis such as Numpy, Pandas, regular expression and scikit-learn. One of the aims of the course is to ensure students are not afraid of python packages and gain ability to freely explore the vast amount of packages python has to offer. The course ends with an exploratory data analysis in machine learning, in which students apply a script-style of programming to describe and understand a dataset of their own choosing. Aside from Python, the course also spends time on several other technologies that are fundamental to the modern practice of data science, including use of the command line and Jupyter notebooks

Learning Outcomes
Upon successful completion of the course, students will be able to:

1. Be able to navigate a file system, manipulate files, and execute programs using a jupyter notebook interface
2. Apply the basic concepts and principles of object-oriented programming including encapsulation, inheritance, and polymorphism.
3. Be able to design, write and test a Python program to implement a working solution to a given problem specification.
4. Know how to use Python to extract data from different type of files and other sources.
5. Know how to read, manipulate, describe, and visualize data using the Numpy and Pandas packages.
6. Know how to leverage Regular Expression / BeautifulSoup for web text extraction and analytics
7. Be able to generate an exploratory analysis of a data set using scikit-learn
8. Be prepared for further programming challenges in more advanced data science courses and to implement statistic and risk management algorithms.

Assessment
Student progress will be assessed through assignments, mid-term projects and final project.

Grading
Final grades will be based on four parts weighted as shown in the table below.

Homework	10%
Midterm Project	40%
Final Project	40%
Attendance/Participation	10%

The weekly assignments are due six days after the corresponding live session. Students will have at least two weeks to work on each group project.

Policies
1. Handing in late homework is strongly discouraged.
2. Cell Phones and other audible devices must be turned off, or set to Silent or Vibrate during class. Students may not text, email, surf the Web, or otherwise use such devices during the class period.
3. Your Instructor is here to help you be successful. If you have questions, concerns or are falling behind, please seek help. It's easier to fall behind than catch up.
4. Ask questions in class if you don't understand.

Academic Dishonesty
The following are common examples of academic dishonesty. This list is not exhaustive.
- Giving your assignment or parts of it to another student. You may not hand in the same code as another student. All of the code that you submit must be written by you.
- Plagiarism. The following are common examples of plagiarism. This list is not exhaustive.
  - Submitting code that you did not write unless that code is from the textbook and you cite it.
  - Text extracted from another source without substantial paraphrasing must be cited as a quote
  - Giving your assignments to another student as a reference point
  - Using another students sign-on (log-on) to complete your work
  - Copying another student's work
  - Submitting another student's work as your own whether it is from this semester or prior semesters or from other classes

Course Schedule (subject to change)

Week	Date	Topic
1	20-Jan	Introduction / Program Level Assessment
2	27-Jan	NumPy
3	3-Feb	Pandas
4	10-Feb	Pandas
5	17-Feb	Pandas
6	24-Feb	Visualization
7	3-Mar	Text Processing
8	10-Mar	Mid Term presentation

9	17-Mar	Spring Break
10	24-Mar	Web Data Gathering
11	31-Mar	Guest Speaker
12	7-Apr	Intro to Scikit-learn
13	14-Apr	Scikit-learn Part II
14	21-Apr	Scikit-learn Part III
15	28-Apr	Final Presentation

HW Schedule - TBD