COURSE SYLLABUS

IST 462/652 SCRIPTING FOR DATA ANALYSIS

Instructor: Carlos E. Caicedo Bastidas
Phone: 315-443-1363
Office: Hinds 215
Email: ccaicedo@syr.edu
Office Hours: TBA
Term: Fall 2022
Course Time: Wednesdays 2:15p.m – 5:00 p.m.
Classroom: Hinds 010

Course Description:

Scripting for data analysis. Acquiring, accessing and transforming data in the forms of structured, semi-structured and unstructured data

Additional Course Description:

The goal of this class is to teach students the tools and skills of scripting needed to solve problems related to the access and preparation of data in a variety of formats and situations, sometimes known as data wrangling. The scripting techniques learned will provide the skills needed to structure data science pipelines, from acquiring and cleaning data to accessing and transforming data for analysis or visualization.

The main content focus will be on information access and processing tasks on structured, semi-structured and unstructured data. For these three types of data, the course will include the use of structured numeric and text data such as that from a spread sheets or databases, the use of data obtained through standard data exchange formats such as HTML or XML from web pages or JSON from web-based APIs, and the use of data obtained by pattern matching from text or log files. The skills learned in this class are intended to complement the analytical and visualization skills learned in other data science courses. Python will be mainly used and taught throughout the course's activities, but it will be assumed that students already have a programming background, either through course work or through online study.

Prerequisite / Co-requisite:

Although no specific course is required, some programming knowledge will be assumed. This may be acquired through courses or online resources

Audience:

Students of the Master's Degree program on Information Management and the Master's Degree program on Applied Data Science are the primary audience for this course. Students in other graduate degree programs may enroll with permission of the instructor.

Credits: 3

Learning Objectives:

After taking this course, the students will be able to:

- Write scripts to access and amass information from files of structured data, access files in semi-structured data and to define and find patterns in unstructured data.
- Prepare and transform data to produce data summaries, lists, and networks.
- Analyze and solve data access problems for the three types of data and to find and deploy appropriate software packages that can be integrated into the problem solution.
- Frame real world data questions and show how they can be answered with data.

Texts / Supplies – Required:

- Python for Data Analysis, 2nd Edition (ISBN-13: 978-1491957660)

Texts / Supplies– Additional:

- Data Wrangling with Python: Tips and Tools to Make Your Life Easier (ISBN-13: 978-1491948811)

Course Requirements and Expectations:

- Homework assignments will be used to re-enforce and evaluate understanding and development of concepts covered in one or two weeks of content.
- After midterms week, students will propose a term project or select from a list provided by the instructor. The instructor must approve the topic/project if proposed by students. The project must involve the development of scripts to achieve a particular data analysis goal or achieve a particular transformation/visualization of the data in a large dataset. The students will prepare a 10 to 20 page project report and deliver a 10 to 15 minute presentation summarizing the project at the end of the semester.
    - Groups of 2 to 4 people will work as a team on a final project depending on the size of the dataset and project objectives. Further details about the project will be provided as the semester progresses.

General Requirement for Assignment Submissions

Assignment submission documents must be professionally prepared with computer applications. Unless otherwise stated, documents must be submitted electronically to Blackboard. No hand-written documents/solutions will be accepted. An assignment submission must be in one document when it is submitted to Blackboard. If you have additional supporting materials that are in physical forms or hard copies (e.g., business forms or some images), you must scan them into JPG or TIFF format and embed them into your MS Word or PDF document. If you use MS Word to prepare your assignments, use Times New Roman style with 12-pt font, 1.5 line spacing, and 1 inch margins.

For each assignment, be sure to include the following information in its first page:

- Course number (i.e. IST 652)
- Name(s) of the author(s) or the team members
- Name of the assignment (e.g. “Homework 2”)
- Date the assignment is due and date in which it is submitted

You must submit all assignments to the professor in class on the deadline specified for each assignment. If you are late, I will deduct 10% of the original grade for the first day of lateness plus 15% for every subsequent day. You will not receive full credit for topics/assignments presented in an unprofessional manner. Professionalism includes the proper use of grammar, punctuation, and limiting spelling mistakes. Professionalism also means adhering to given instructions. Failure to adhere to the assignment instructions will result in a reduction of the grade. If English is not your first language set up an appointment with the writing program so they can help you improve your writing.

How to Succeed in This Course:
Keep pace with the content of the course. It is expected that you will spend 6 to 9 hours outside of class time covering the material, the readings and executing labs / homework and project assignments.

Mental health and overall well-being are significant predictors of academic success. As such it is essential that during your college experience you develop the skills and resources to effectively navigate stress, anxiety, depression and other mental health concerns. Please familiarize yourself with the range of resources the Barnes Center provides (https://ese.syr.edu/bewell/) and seek out support for mental health concerns as needed. Counseling services are available 24/7, 365 days, at 315-443-8000. I encourage you to explore the resources available through the Wellness Leadership Institute, https://ese.syr.edu/bewell/wellness-leadership-institute/.

Grading:

All exams and assignments will be graded in a 0-100 point scale. The weight distribution of the different activities/assignments that you will work on in this course is mentioned in the following table.

Activity / Assignment	Final grade percentage value
Homework & quizzes	40%
Attendance and participation in class sessions/activities	10%
Midterm exam	12%
Final project & presentation	23%
Final exam	15%

Grading Table

The following table will be used in the assignment of your final letter grade for the course.

Letter Grade	Grade points /credit	Percentage Range	Total Points
A	4.000	93% - 100%	93 – 100
A-	3.667	90% - 92.99%	90 – 92.99
B+	3.333	87% - 89.99%	87 – 89.99
B	3.000	83% - 86.99%	83 – 86.99
B-	2.667	80% - 82.99%	80 – 82.99
C+	2.333	77% - 79.99%	77 – 79.99
C	2.000	73% - 76.99%	73 – 76.99
C-	1.6667	70% - 72.99%	70 – 72.99
F	0	0% - 69.99%	0 – 69.99

Syracuse University Policies:

Syracuse University has a variety of policies designed to guarantee that students live and study in a community respectful of their needs and those of fellow students. **The policies and services are listed on the new Syracuse University Senate approved syllabus appendix titled, 'Syracuse University Student Policies and Services'. These statements are an official part of this course syllabus.**

Course specific policies

Attendance

In addition to the university's attendance policy, please note that attendance will be taken throughout the semester in both labs and lectures. If you arrive late or leave early, you will be marked absent. There are no excused absences unless documented by the university. If you have 2 or more absences, your final grade will be dropped one level down the grade scale. (A- is lowered to B+, C+ becomes a C, etc.)

Use of Class Materials and Recordings:

Original class materials (handouts, assignments, tests, etc.) and recordings of class sessions are the intellectual property of the course instructor. You may download these materials for your use in this class. However, you may not provide these materials to other parties (e.g., web sites, social media, other students) without permission. Doing so is a violation of intellectual property law and of the student code of conduct.

Educational Use of Student Work

Student work prepared for University courses in any media may be used for educational purposes, if the course syllabus makes clear that such use may occur. You grant permission to have your work used in this manner by registering for, and by continuing

to be enrolled in, courses where such use of student work is announced in the course syllabus.

I intend to use academic work that you complete this semester in subsequent semesters for educational purposes. Before using your work for that purpose, I will either get your written permission or render the work anonymous by removing all your personal identification.

Course Schedule:

The following table shows the breakdown of topics to be covered in this course. This table is **subject to modification as the course evolves**. Up to date information (especially for homework assignments) will be posted in Blackboard.

Week #	Date	Topic
1	8/31	Introduction. Data pipeline and Python foundations
2	9/7	Python Basics 1 - Data pipeline and Python foundations - Working with Jupyter Notebooks
3	9/14	Python Basics 2 Working with NumPy
4	9/21	Structured Data Working with Numpy 2 Working with Pandas
5	9/28	Structured Data Working with Pandas 2
6	10/5	Working with Semi-structured Data - JSON - Working with APIs
7	10/12	**Midterm Exam** Web-scrapping 1
8	10/19	Web-scrapping 2 Working with databases
9	10/26	Unstructured Data Data cleaning
10	11/2	Time series
11	11/9	Advanced pandas
12	11/16	**THANKSGIVING**
13	11/23	Data analytics tools Advanced pandas 2
14	11/30	Final project presentations
15	12/7	Final project presentations **Final exam (online)**