

---

## References

There are no sources in the current document.

## Description

Scripting for the data science pipeline. Acquiring, accessing, and transforming data in the forms of structured, semistructured, 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 of accessing and preparing data in a variety of formats and situations, sometimes known as *data wrangling*. The scripting will provide the skills needed to form data science pipelines, from acquiring and cleaning data to accessing data and transforming data for analysis or visualization.

The main content focus is on information access and processing tasks on the types of structured, semistructured, and unstructured data in current use in information applications. For these three types of data, the course includes the use of structured numeric and text data such as that from a spreadsheet or database, 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 scripting language Python was chosen because of its ease of use and available packages to work with data in many information applications. The skills learned in this class are intended to complement the analytical and visualization skills learned in other data science courses. The scripting language Python will be taught, but it will be assumed that students already have a programming background, through either coursework or online study.

## Prerequisite/Corequisite

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

## Credit

3.0

## Learning Objectives

Upon successful completion of this course, the student will be able to:

1. Write scripts to access and amass data from fields in structured data, access fields in semistructured data, and define and find patterns of data in unstructured data
2. Prepare and transform data to produce data summaries, lists, and networks
3. 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
4. Frame real-world data questions and show how they can be answered from data

---

### **Bibliography/Texts/Supplies (Required)**

*Python for Data Analysis*: <https://www.programmer-books.com/wp-content/uploads/2019/04/Python-for-Data-Analysis-2nd-Edition.pdf> (downloadable PDF)  
Available free online (downloadable PDF) or for purchase as a (paperback) book through Amazon or other sites.

Other readings will be assigned from online resources.

### **Supplemental Text (Not Required)**

Bader, D. (2017). *Python Tricks: The Book*. Dan Bader.  
Myers, M. (2017). *A Smarter Way to Learn Python* (1st ed.). Mark Myers. Retrieved from <https://www.ASmarterWayToLearn.com>  
Nelson, D. (2020). *Data Visualization in Python*. StackAbuse.com.

### **Course Requirements and Expectations**

Class sessions are designed around developing solutions for the main focus application tasks. Each section of the course contains a smaller number of daily problems that lead up to focus on one overall application design challenge in each of the three types of data. The instructor will then demonstrate how to solve an example of the design challenge, and then students will be asked to solve a similar problem. The design challenges will focus on the use of structured numeric and text data such as that from a spreadsheet or database, 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 coursework will consist mainly of small weekly lab programming and analysis problems, two larger design and programming problems for homework, and the final project. There will also be two quizzes, where half the quiz will be multiple-choice questions and the other half will be programming questions. The class attendance and participation grade include both asynchronous (completing asynchronous exercises) AND live session attendance and participation. The final project will require an additional report on a data science analysis question appropriate for the problem data.

COURSE SYLLABUS  
**IST 652 Scripting for Data Analysis**

3

---

<b>5 lab exercises</b> ( <i>grading scale is 100 points</i> )	(5 x 4% each) = 20%
<b>2 quizzes</b> ( <i>grading scale is 100 points</i> )	(2 x 10% each) = 20%
<b>Participation</b>	(5% asynchronous + 5% live sessions) = 10%
<b>2 homework assignments</b> ( <i>grading scale is 100 points</i> )	(2 x 15% each) = 30%
<b>Final project</b> ( <i>grading scale is 100 points for each   grading breakdown: proposal 2%, presentation 3%, report and program 15%</i> )	20%
<b>Total</b>	100%

### **Grading**

Grades	Grade Points/ Credit	Percentage Range
A	4.00	94–100.0
A–	3.66	90–93.9
B+	3.33	87–89.9
B	3.00	84–86.9
B–	2.66	80–83.9
C+	2.33	76–79.9
C	2.00	73–75.9
C–	1.66	70–72.9
F	0.00	below 70

### **University Attendance Policy**

Attendance in classes is expected in all courses at Syracuse University. Students are expected to arrive on campus in time to attend the first meeting of all classes for which they are registered. Students who do not attend classes starting with the first scheduled meeting may be academically withdrawn as not making progress toward degree by failure to attend. Instructors set course-specific policies for absences from scheduled class meetings in their syllabi. It is a federal requirement that students who do not attend or cease to attend a class to be reported at the time of determination by the faculty. Students should also review the university's religious observance policy and make the required arrangements at the beginning of each semester.

### **iSchool Values**

The culture, curriculum, and decision making at the iSchool are guided by our shared values.

### **Excellence**

We are a student-centered institution committed to learning and intellectual diversity. As a community of scholars and practitioners, we pursue research and teaching excellence through the development, integration, and application of knowledge.

---

### **Discovery and Innovation**

We value discovery and innovation to advance the information fields. We identify new opportunities, respond creatively to emerging issues, and lead our fields through partnerships and learning communities.

### **Integrity**

As information professionals, we uphold intellectual honesty and responsibility. We carefully consider the implications of our actions, taking fairness and equity into consideration. We contribute to, and promote, the highest standards for the ethical use of information and technology.

### **Diversity and Inclusion**

We strive for diversity in our community and celebrate difference. We embrace a multiplicity of voices to address social and technical challenges through interdisciplinary analysis and solutions.

### **Global Citizenship and Engagement**

We enable our iSchool community to participate actively as global citizens working to advance the common good. We have a responsibility to inform and improve society, and to influence policy through pedagogy, research, and advocacy.

### **Course-Specific Policies on Attendance, Late Work, and Makeup Work**

Attendance is required and cannot be made up except for excused absences for illness and for other school-sanctioned activities. For assignments, late work will be accepted but will be penalized with the points equivalent to ½ of a letter grade for one day late and on a sliding scale for additional lateness of up to a full letter grade.

### **Syracuse University Policies**

Students should review the university's policies regarding:

**Diversity and Disability:** <https://www.syracuse.edu/life/accessibilitydiversity/>

**Religious Observances Notification and Policy:**

[http://supolicies.syr.edu/studs/religious\\_observance.htm](http://supolicies.syr.edu/studs/religious_observance.htm)

**Orange SUccess:** <http://orangesuccess.syr.edu/getting-started-2/>.

### **Disability-Related Accommodations**

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. If you believe that you need academic adjustments (accommodations) for a disability, please contact the Office of Disability Services (ODS), visit the ODS website (<http://disabilityservices.syr.edu>), located in Room 309 of 804 University Avenue), or call (315) 443-4498 or TDD: (315) 443-1371 for an appointment to discuss your needs and the process for requesting academic adjustments. ODS is responsible for coordinating disability-related academic adjustments and will issue students with documented

---

Disabilities Accommodation Authorization Letters, as appropriate. Since academic adjustments may require early planning and generally are not provided retroactively, please contact ODS as soon as possible. Our goal at the iSchool is to create learning environments that are useable, equitable, inclusive, and welcoming. If there are aspects of the instruction or design of this course that result in barriers to your inclusion or accurate assessment or achievement, please meet with me to discuss additional strategies beyond official accommodations that may be helpful to your success.

### **Academic Integrity Policy**

Syracuse University's Academic Integrity Policy reflects the high value that we, as a university community, place on honesty in academic work. The policy defines our expectations for academic honesty and holds students accountable for the integrity of all work they submit. Students should understand that it is their responsibility to learn about course-specific expectations, as well as about university-wide academic integrity expectations. The policy governs appropriate citation and use of sources, the integrity of work submitted in exams and assignments, and the veracity of signatures on attendance sheets and other verification of participation in class activities. The policy also prohibits students from submitting the same work in more than one class without receiving written authorization in advance from both instructors. Under the policy, students found in violation are subject to grade sanctions determined by the course instructor and non-grade sanctions determined by the School or College where the course is offered as described in the Violation and Sanction Classification Rubric. SU students are required to read an online summary of the University's academic integrity expectations and provide an electronic signature agreeing to abide by them twice a year during pre-term check-in on MySlice. For more information about the policy, see <https://class.syr.edu/academic-integrity/policy/>

The Violation and Sanction Classification Rubric establishes recommended guidelines for the determination of grade penalties by faculty and instructors, while also giving them discretion to select the grade penalty they believe most suitable, including course failure, regardless of violation level. Any established violation in this course may result in course failure regardless of violation level.

### **Course-Specific Policy on the Use of Turnitin**

This class will use the plagiarism detection and prevention system Turnitin. I will submit all papers you write for this class to Turnitin, which compares submitted documents against documents on the internet and against student papers submitted to Turnitin at SU and at other colleges and universities. I will take your knowledge of the subject matter of this course and your writing level and style into account in interpreting the originality report. Keep in mind that all papers you submit for this class will become part of the Turnitin.com reference database solely for the purpose of detecting plagiarism of such papers. I will provide you a consent statement to sign, date, and return to me that will state the following:

"By signing below, I give permission for the submission of all work I turn in for *Scripting for Data Analysis* to the plagiarism detection and prevention system

---

Turnitin, which compares submitted documents against documents on the internet and against student papers submitted to Turnitin at SU and at other colleges and universities. I understand that all assignments submitted for this class will become part of the Turnitin.com reference database solely for the purpose of detecting plagiarism of such papers.”

### **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.

### **Discrimination or Harassment**

The University does not discriminate and prohibits harassment or discrimination related to any protected category including creed, ethnicity, citizenship, sexual orientation, national origin, sex, gender, pregnancy, disability, marital status, age, race, color, veteran status, military status, religion, sexual orientation, domestic violence status, genetic information, gender identity, gender expression, or perceived gender.

Any complaint of discrimination or harassment related to any of these protected bases should be reported to Sheila Johnson-Willis, the University's Chief Equal Opportunity & Title IX Officer. She is responsible for coordinating compliance efforts under various laws, including Titles VI, VII, IX and Section 504 of the Rehabilitation Act. She can be contacted at Equal Opportunity, Inclusion, and Resolution Services, 005 Steele Hall, Syracuse University, Syracuse, NY 13244-1120; by email: [titleix@syr.edu](mailto:titleix@syr.edu); or by telephone: 315-443-0211.

Federal and state law, and University policy, prohibit discrimination and harassment based on sex or gender (including sexual harassment, sexual assault, domestic/dating violence, stalking, sexual exploitation, and retaliation). If a student has been harassed or assaulted, they can obtain confidential counseling support, 24 hours a day, 7 days a week, from the Sexual and Relationship Violence Response Team at the Counseling Center (315-443-4715, 200 Walnut Place, Syracuse, New York 13244-5040). Incidents of sexual violence or harassment can be reported non-confidentially to the University's Title IX Officer (Sheila Johnson Willis, 315-443-0211, [titleix@syr.edu](mailto:titleix@syr.edu), 005 Steele Hall). Reports to law enforcement can be made to the University's Department of Public Safety (315-443-2224, 005 Sims Hall), the Syracuse Police Department (511 South State Street, Syracuse, New York; 911 in case of emergency; or 315-435-3016 to speak with the Abused Persons Unit), or the State Police (844-845-7269). I will seek to keep information you share with me private to the greatest extent possible, but as a professor I have mandatory reporting responsibilities to share information regarding sexual misconduct, harassment, and crimes I learn about to help make our campus a safer place for all.

---

### **School Library Media Program Assessment**

The School Library Media Program is accredited by the Council for the Accreditation of Educator Preparation (CAEP) through the Syracuse University School of Education. As a part of that accreditation, the School Library Media Program must assess student performance on the competencies that correlate to program outcomes. The competencies which are assessed are identical to the items on your Competency Checklist. As a part of School Library Media Program planning, course-embedded assessments have been aligned with student competencies. For CAEP reporting, each faculty member with competency-based, course-embedded assessments is asked to rate (1=Ineffective, 2=Developing, 3=Effective, or 4=Highly Effective) candidates' performance on the respective competencies. This is the same rating scale students use when completing the competency checklist. Scoring is conducted for key assignments and not all assignments for a course.

What this means for you:

Your individual score is NOT a grade and it is part of an aggregate report. If a student is performing at an Ineffective or Developing level, a comment is submitted with the score, which is also aggregated. Individual scores and comments are not associated with specific student names.

### **Course Evaluations**

There will be an end-of-course evaluation for you to complete this term. This evaluation will be conducted online and is entirely anonymous. You will receive a notification from the Syracuse University Office of Institutional Research & Assessment (OIRA) department in your email account with the evaluation website link and your passcode. Please take the time and fill out this evaluation as your feedback and support of this assessment effort is very much appreciated. The school carefully reviews ratings and comments that you submit, and these factor into decisions about course, program, and instructor development.

**Course Schedule:** Week/lecture, topic for the week/lecture, and required readings are in the table below.

COURSE SYLLABUS  
IST 652 Scripting for Data Analysis

8

Week/ Lecture	Topic	Required Reading and Assignment
1	Data Pipeline and Python Language Basics	<b>Required Readings</b> <i>Python for Data Analysis</i> , Chapters 1 and 2 <a href="#">Data Science: A Kaggle Walkthrough—Introduction</a>
2	Booleans and Dictionaries	<b>Required Readings</b> <i>Python for Data Analysis</i> , Chapter 3 <a href="#">Data Exploration: A Kaggle Walkthrough—Understanding the Data</a> <b>Assignment Due</b> Lab Exercise 1
3	Exploring and Transforming Data for Structured Data	<b>Required Readings</b> <i>Python for Data Analysis</i> , Chapters 5 and 6 <b>Quiz 1 will open after completion of asynchronous.</b> <b>Assignment Due</b> Lab Exercise 2
4	Arrays, Functions, and Categorical Summarization	<b>Required Readings</b> <i>Python for Data Analysis</i> , Chapters 4 and 10 <a href="#">Cleaning the Data: A Kaggle Walkthrough</a> <a href="#">Python NumPy Tutorial</a> <b>Assignments Due</b> Lab Exercise 3 Quiz 1 must be completed by the end of Week 4.



COURSE SYLLABUS  
IST 652 Scripting for Data Analysis

9

Week/ Lecture	Topic	Required Reading and Assignment
5	Stacking and Unstacking Data	<b>Required Readings</b> <i>Python for Data Analysis</i> , Chapters 8 and 11 <a href="#">Pandas: How to Pivot Data</a> <a href="#">Pandas: Joining Tables</a> <a href="#">Data Science: A Kaggle Walkthrough—Transforming the Data</a> <b>Assignment Due</b> Homework 1
6	Semistructured Data	<b>Required Readings</b> <i>Python for Data Analysis</i> : Chapter 9 <a href="#">MDN Web Docs: HTML</a> <a href="#">W3Schools HTML Element Reference</a> <a href="#">Dominos: Data-Driven Decision Making at the World's Largest Pizza Delivery Chain</a> <b>Assignments Due</b> Lab Exercise 4 Final Project Proposal
7	Mongo Database   JSON From APIs	<b>Required Readings</b> <a href="#">NoSQL Databases: An Overview</a> <a href="#">PyMongo Tutorial</a> <a href="#">JSON Encoder and Decoder</a> <a href="#">XML vs. JSON</a> <b>Quiz 2 will open after completion of asynchronous.</b> <b>Assignment Due</b> Lab Exercise 5

COURSE SYLLABUS  
IST 652 Scripting for Data Analysis

10

Week/ Lecture	Topic	Required Reading and Assignment
8	Processing Twitter and Facebook	<b>Required Readings</b> <i>Python for Data Analysis</i> , Chapters 12 and 13 <a href="#">Data Science: A Kaggle Walkthrough—Adding New Data</a> <a href="#">12 Twitter Sentiment Analysis Algorithms Compared</a> <b>Assignments Due</b> Homework 2 Quiz 2 must be completed by the end of Week 8.
9	Unstructured Data	<b>Required Readings</b> <i>Python for Data Analysis</i> , Chapter 14 <a href="#">Regular Expression Documentation</a> <a href="#">Data Science: A Kaggle Walkthrough—Creating a Model</a>
10	Network Structures	<b>Required Readings</b> <a href="#">How to Go Viral on Twitter</a> (10 Easy Strategies) <a href="#">Python NetworkX Package Tutorial</a> <b>Assignments Due</b> Final Project Report and Program Final Project Presentations
11	Presentations	<b>Final Project Presentations</b>

---

## **Lab Exercises**

Directions for each of the lab exercises and any needed data files are provided in the Toolbox of the course.

- 1: DUE Friday by 11:59 p.m. Eastern Time at the end of Week 2
- 2: DUE Friday by 11:59 p.m. Eastern Time at the end of Week 3
- 3: DUE Friday by 11:59 p.m. Eastern Time at the end of Week 4
- 4: DUE Friday by 11:59 p.m. Eastern Time at the end of Week 6
- 5: DUE Friday by 11:59 p.m. Eastern Time at the end of Week 7

## **Quizzes**

Quiz 1:

Opens upon completion of asynchronous activities in Week 3  
DUE: Friday by 11:59 p.m. Eastern Time at the end of Week 4

Quiz 2:

Opens upon completion of asynchronous activities in Week 7  
DUE: Friday by 11:59 p.m. Eastern Time at the end of Week 8

## **Homework 1: Structured Data**

DUE Friday by 11:59 p.m. Eastern Time at the end of Week 5

## **Homework 2: Semistructured Data**

DUE Friday by 11:59 p.m. Eastern Time at the end of Week 8

---

## **Final Project**

Final Project Proposal

DUE Friday by 11:59 p.m. Eastern Time at the end of Week 6

Final Project Report and Program

DUE Friday by 11:59 p.m. Eastern Time at the end of Week 10

Final Project Presentations

To be presented in Weeks 10 and 11 live sessions