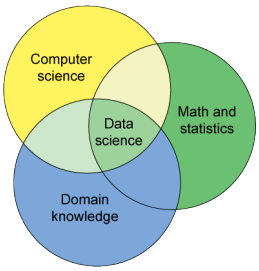
Introduction to Data Science

CMPSC-310

Spring 2020



Instructor Information

Instructor	Dmitry Zinoviev
E-mail	dzinoviev@suffolk.edu
Phone	x1985 (e-mail always works faster)
Office and Office Hours	TTh 4:30 PM–5:30 PM, M 1:30 PM–2:30 PM or by appointment TR73-8080

Course Information

Meeting Times	TTh 9:25AM–10:40AM
Location	SAM-318
Catalog Description	The field of data science is emerging at the intersection of the fields of social science and statistics, information and computer science, and design. Data science involves using automated methods to analyze massive amounts of data and to extract knowledge from them. This course serves as a project-based introduction to data science in Python, covering data organization, retrieval, statistical data processing, and data visualization.
Prerequisites	CMPSC-F132 (in Python; required) or permission of instructor. MATH-134, MATH-165, MATH-255, or STAT-250 (at least one course required).
Credit Hours	Four credit hours. This course follows the Federal Government's Credit Hour definition. For more information regarding the definition, please see the Suffolk University Syllabus webpage: http://www.suffolk.edu/syllabus.
Textbook / Course Materials	D. Zinoviev, "Data Science Essentials in Python," Pragmatic Programmer, 2016; recommended; can be obtained in the university bookstore
Software	You will need access to a Linux, Windows, or macOS computer with working Python (version 3.4 or above). Consider using the free Anaconda distribution. You can download Anaconda from https://store.continuum.io/cshop/anaconda/ (Choose Anaconda for Python 3.4 or above, not 2.7.)

Course Goals and Learning Objectives

Goals	Objectives	Assessments
Upon successful completion of this course, students should be able to know/understand:	**Upon successful completion of this course, students should be able to:**	**How the student will be assessed on these learning objectives:**
...how to recognize, extract, organize, and interpret quantitative data by incorporating best practices in data management.	...read and understand quantitative data in various formats.	PR1–4; MT; final
	...communicate the meaning of quantitative data and the results of data analysis.	PR1, 2, 4
	...develop software that reads and normalizes plain-text and HTML data and reads structured data (JSON, CSV).	PR1–4; MT; final
...the theoretical basis of quantitative reasoning, how to process data by logically manipulating and joining the underlying raw data into a new representation and packaging.	...explain the basic concepts of quantitative reasoning, such as variables, constants, and estimates.	PR4; MT; final
	...develop software that calculates data statistics and dependencies (such as frequency analysis; distributions; correlations; regressions; complex data networks), using *NumPy*, *networkx*, and other Python modules	PR2–4
	...understand how inferences are drawn from quantitative analysis.	PR2, 4; final
	...recognize the limitations of quantitative methods.	PR4
...the practical application of quantitative data analysis, how to deliver data ensuring that the message the data has is being accessed by those that need to hear it.	...determine and use appropriate quantitative methods (including visualization methods for scalar, vector, and network data) to solve problems.	PR3, 4; final
	...accurately interpret the results of data analysis.	PR2, 4; final
	...assess results for reasonableness.	PR2, 4

Course Schedule

Week of	Topic	Assessment
1/13	Review of Python data structures, file and string methods	Entry quiz
1/20	Review of Python data structures, file and string methods	
1/27	Structured text data formats (CSV, HTML, JSON)	
2/03	Structured text data formats (CSV, HTML, JSON) Regular expressions *2/04. The last day students can withdraw from a course without receiving a grade of **W***	
2/10	Regular expressions Text analysis (*not on the midterm!*)	
2/17	*No class on 2/18! ("Suffolk Monday.")*	Midterm
2/24	Text analysis	
3/02	Numeric data. *NumPy*	
~~3/09~~	*Spring break*	
3/16	Data frames. *Pandas* *3/20. The last day students can withdraw from a course without receiving a grade of **F***	
3/23	Data frames. *Pandas*	
3/30	Statistics for data science, correlations and regressions	
4/06	Intro to machine learning	
4/13	Network data	
4/20	Building and analyzing data networks with *networkx* *4/23 Last class*	
4/29	Final exam in SAM-318 at 2:00 PM–4:30 PM (150 minutes)	

Continuity of Learning Plan

In case of snow days or other course cancellations, go to the section "Syllabus/Snow Emergency Assignments" on BlackBoard and follow the instructions.

Assignments/Exams/Papers/Projects

Students will be evaluated in the following areas.

Quizzes

Approximately weekly multiple-choice quizzes will be posted on BlackBoard.

Programming Homework

Your programming homework consists of at least four team programming projects and at least five individual programming exercises. The complete solutions must be submitted to BlackBoard.

You will work on the programming projects in teams of two or three. You are responsible for your homework. All programming homework must be done independently or in the pre-assigned teams. A close similarity of homework papers or programs submitted by different students is a sufficient cause not to grade all of them.

There is a penalty of 25% per late day for all late programming assignments unless a severe medical or family condition prevents you from submitting them on time.

Exams

There will be one 75-minute midterm exam and a 150-minute final exam. All exams are on paper, closed book, closed notes. You are allowed to bring a two-sided "cheat sheet." I do not usually allow make-up exams.

Select assignments in this course may be used by our accreditation team for institutional assessment purposes and will be handled confidentially.

Grading and Evaluation

The estimation of your course grade is calculated using the following weights:

- Quizzes (multiple choice)—10%.
- Programming projects—25%.
- Programming exercises—15%.
- Midterm exam (on paper)—25%.
- Final exam (on paper)—25%.

Your grades will be posted at BlackBoard. Please consider the online grade book as a courtesy to you, subject to errors given various upgrades and shifts in the software. I reserve the right to make grade book corrections to keep it consistent with the syllabus so that your grade reflects actual performance, not software or user error.

Student Engagement Hours

To complete this course, students will need to dedicate, at a minimum, the following amount of time to the listed activities:

Assignment/Activity	Engagement Estimate	Engagement Hours
Programming exercises	3 hours each × 5 exercises	15
Programming projects	14 hours each × 4 projects	56
Class attendance	3 hours × 14 weeks	42
Course readings	200 pages × 10 minutes per page	33
Final exam preparation	11 hours	11
Lecture notes review	60 pages × 10 minutes per page	10
Midterm exam preparation	5 hours	5
Take-home quizzes	1.5 hour each × 7 quizzes	10.5
Total		182.5

Course Policies

Missing Classes

Being in class and participating in discussion and problem solving is an essential element of the course curriculum. In case you need to miss a class, as a courtesy, please send an email informing me about your situation. If you miss more than four meetings, you final grade may be reduced by as much as 25%.

Being Late

Late students disrupt the normal class flow and deflect my and other students' attention from the subject. Students systematically coming five or more minutes late may have their final grade reduced by as much as 10%.

Suffolk University Syllabus Policies

This course adheres to policies and procedures that apply to all Suffolk courses concerning disability accommodation, academic misconduct, academic grievance, attendance, and credit hour compliance. The university policies can be found here: www.suffolk.edu/syllabus.

Student Resources

The university provides a range of academic, counseling, medical and administrative student support services. To learn more, explore this webpage: www.suffolk.edu/syllabus.