MSDS 430 Syllabus

Instructor Contact Information

Instructor: Dr. Moses Boudourides

Email: moses.boudourides@northwestern.edu I will respond to emails within 24 – 48 hours.

Phone: +30-6948886094 During normal business hours (ET); Leave a message if no answer; Please no texts.

Links: http://mboudour.github.io/about/

Office Hours: By appointment

Optional Synchronous Meetings: TBA

Biography

Dr. Boudourides was Professor of Mathematics at the University of Patras in Greece (from 1998 until September 2017, when he retired). In 2019-20, he was Visiting Professor at the Science Division of the New York University in Abu Dhabi. Previously, he was Associate Professor at the Department of Electrical and Computer Engineering of the Democritus University of Thrace in Xanthi, Greece (1982-1998). In addition, he was Visiting Professor at the Department of Mathematics of the University of California Irvine (1990-1991). His PhD is from the Johns Hopkins University (1978). His undergraduate studies were at the National Technical University of Athens in Greece, from where he graduated with a Diploma in Chemical Engineering (1973). His research interests and publications are on dynamical systems, social network data science, computational social science and digital humanities. In 2019 he was awarded a Robert K Merton Visiting Research Fellowship from the Institute for Analytical Sociology (IAS) at Linköping University in Sweden.

Teaching Assistant Contact Information

Teaching Assistant: Geetha Ramasamy

Email: geetharamasamy2019@u.northwestern.edu Messaging via Canvas is the best way to reach Geetha. She will respond to queries within 24 – 48 hours.

Biography

Geetha works as a data scientist on the Advanced Analytics team in health insurance space where she applies disciplined analytics to predict behavior and optimize programs and products to maximize revenue growth. She has experience in using data science skills to see sizeable dollar impact on business as well as determining the impact on Enterprise wide goal setting KPIs. She uses python at work and leverage the use of pyspark for big data applications. She graduated from the same program in March 2018 when it was still named MS in predictive Analytics. She has worked in various domains including telecom, manufacturing, insurance and credit cards, developing windows and web applications. Geetha has been a Teaching Assistant in Database Systems and Data Preparation and Practical Machine Learning. She lives in the suburbs of Chicago with her husband and two kids.

Course Description

This course introduces core features of the Python programming language, demonstrating fundamental concepts in computer science. It provides an in-depth discussion of data representation strategies, showing how data structures are implemented in Python and demonstrating tools for data science and software engineering. Working on data analysis problems, students employ various programming paradigms, including functional programming, object-oriented programming, and data stream processing. Special attention is paid to the standard Python library and packages for analytics and modeling.

Course Objectives

By the end of this course, you will be able to:

- Apply basic Python programming techniques
- Perform a variety of mathematical and statistical calculations using Python
- Recognize data structures and control structures to build Python code
- Execute Python code containing libraries commonly used in data science
- Manipulate data files and internet data using Python
- Build, organize, and generalize Python code for reusability

---

Prerequisites

There are no prerequisites for this course.

Diversity Statement

I truly believe that we are all unique individuals and each of us adds value to the world around us. We all deserve respect and equity regardless of our backgrounds or identities. I embrace the knowledge that I have gained from my students over the years and look forward to learning more from you. Your opinions, approaches, knowledge, and experiences are important to me and I welcome broadening my experiences and knowledge by working with you. I teach with the approach that I am more than simply a subject matter expert. I am here to not only guide you, but to work alongside you as you take this journey through this course.

Required and Optional Readings and Resources

Required Readings and Resources

- How to Think Like a Computer Scientist: Interactive Edition, an open source interactive text based on *How to Think Like a Computer Scientist* [A. Downey, J. Elkner, C. Meyers].
- The Python Tutorial – Documentation available from www.python.org.
- LinkedIn Learning – *Python Essential Training* [B. Weinman]
- O'Reilly Books Online – *Become a Python Data Analyst* [A. Fuentes]
- O'Reilly Books Online – *Python Programming Language* [D. Beazley]; *Become a Python Data Analyst* [A. Fuentes]

Assigned readings and required media are posted on Canvas.

Optional Readings and Resources

- Beazley, D. & Jones, B. K. 2013. *Python Cookbook, 3rd Ed*. Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-1-449-34037-7]
- Downey, A. B. (2015). *Think Python, 2nd Ed.* Sebastopaol, Calif.: O'Reilly. [ISBN-13: 9781491939369] (No purchase necessary. Available as a free electronic download at Think Python or at Amazon for purchase as either an eBook or traditional text.)
- VanderPlas, J., 2017. *Python Data Science Handbook: Essential Tools for Working with Data*. Sebastopol, Calif.: O'Reilly [ISBN-13: 978-1491912058]
- LinkedIn Learning – *Learning Python* [J. Marini]

Assignment Overview and Grading Breakdown

Grading and feedback turnaround will be no more than one week from the due date. You will be notified if turnaround will be longer than one week.

The discussion forums will be graded based on specific criteria or a rubric. The criteria or rubric for each discussion will be available in the course. To view the discussion forum rubric, click the gear icon in the upper right corner of the page and choose Show Rubric.

Module	Assignment	Description	Percentage
Modules 1 – 10	**Discussions**	Weekly class participation is determined by posting to discussion forums. You will be expected to provide quality insights in class discussion forums in a timely fashion and interact with others in a manner that provides depth and insight into the discussion. Initial discussion posts based on the prompt provided will be due by 11:59pm CT on Thursday of each week and at least two additional posts to peers will be due by 11:59pm CT on Sunday of each week.	10%
Modules 1 – 9	**Python Assignments**	Weekly assignments are designed to give students experience using tools from the readings and resources will be completed using Jupyter Notebook. These assignments will be due by 11:59pm CT on Sunday of each week. Point values for each Python assignment will vary.	50%

Module 4	**Project Proposal**	Students will select data and outline plans to prepare for their EDA project.	5%
Module 6	**EDA Phase 1**	Students will complete the first phase of an EDA project by setting goals, retrieving data, cleaning data, and transforming data. Students have the option to work on the EDA project in groups of 2 or 3 or individually.	10%
Module 8	**EDA Phase 2**	Students will complete the next phase of their EDA project by creating visualizations, analyzing results, and forming conclusions.	15%
Module 10	**EDA Report**	Students will compile the processes and results from the first two phases of the EDA project into a managerial style report.	10%
		Total	100%

Grading Scale

Grade	Percentages
A	93-100
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
F	Below 70

The School of Professional Studies does not award D grades in graduate coursework.

Late Work Policy

Unless otherwise noted, all work is due on the assigned day by 11:59 PM (Central Time). This includes exams and participation in the discussions. Late work is not accepted unless prior arrangements have been made with the instructor, in which case late penalties may apply. Discussion posts after the week has ended will not be accepted.

Online Communication and Interaction Expectations

Discussion Forums

The purpose of the discussion boards is to allow students to freely exchange ideas. It is imperative to remain respectful of all viewpoints and positions and, when necessary, agree to respectfully disagree. While active and frequent participation is encouraged, cluttering a discussion board with inappropriate, irrelevant, or insignificant material will not earn additional points and may result in receiving less than full credit. Frequency matters but contributing content that adds value is paramount. Please remember to cite all sources—when relevant—in order to avoid plagiarism. Please post your viewpoints first and then discuss others' viewpoints.

The quality of your posts and how others view and respond to them are the most valued. A single statement mostly implying "I agree" or "I do not agree" is not counted as a post. Explain, clarify, politely ask for details, provide details, persuade, and enrich communications for a great discussion experience. Please note, there is a requirement to respond to at least two fellow class members posts. Also, remember to cite all sources—when relevant—in order to avoid plagiarism.

All initial posts addressing the prompt are due by 11:59 CT on Thursdays each week and there is a requirement to respond to at least two fellow class members' posts. The detail of what is expected is available in the discussion rubric.

Online Communication Etiquette

Beyond interacting with your instructor and peers in discussions, you will be expected to communicate by Canvas message, email, and sync session. Your instructor may also make themselves available by phone or text. In all contexts, keep your communication professional and respect the instructor's posted availability. To learn more about professional communication, please review the Communicating Effectively with Faculty guide.

Just as you expect a response when you send a message to your instructor, please respond promptly when your instructor contacts you. Your instructor will expect a response within two business days. This will require that you log into the course site regularly and set up your notifications to inform you when the instructor posts an announcement, provides feedback on work, or sends you a Canvas message. For guidance on setting your notifications, please review How do I set my Canvas notification settings as a student? It is also recommended that

you check your u.northwestern e-mail account regularly, or forward your u.northwestern e-mail to an account you check frequently.

Participation and Attendance

This course will not meet at a particular time each week. All course goals, session learning objectives, and assessments are supported through classroom elements that can be accessed at any time. To measure class participation (or attendance), your participation in discussion boards is required, graded, and paramount to your success in this course. Please note that any scheduled synchronous meetings are optional. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation.

Student Support Services

AccessibleNU

This course is designed to be welcoming to, accessible to, and usable by everyone, including students who are English-language learners, have a variety of learning styles, have disabilities, or are new to online learning. Be sure to let me know immediately if you encounter a required element or resource in the course that is not accessible to you. Also, let me know of changes I can make to the course so that it is more welcoming to, accessible to, or usable by students who take this course in the future.

Northwestern University and AccessibleNU are committed to providing a supportive and challenging environment for all undergraduate, graduate, professional school, and professional studies students with disabilities who attend the University. Additionally, the University and AccessibleNU work to provide students with disabilities and other conditions requiring accommodation a learning and community environment that affords them full participation, equal access, and reasonable accommodation. The majority of accommodations, services, and auxiliary aids provided to eligible students are coordinated by AccessibleNU, which is part of the Dean of Students Office.

SPS Student Services

The Department of Student Services supports the academic and professional growth of SPS students. The Student Services team guides students through academic planning, policies, and administrative procedures, and promotes a supportive environment to foster student success. Students are encouraged to actively make use of the resources and staff available to assist them: Academic and Career Advisers, Counseling and Health Services, Student Affairs, Legal Services, Financial Aid and Student Accounts, among other services.

For a comprehensive overview of course and program processes and policies and helpful student resources, please refer to your SPS Student Handbook.

Academic Support Services

Northwestern University Library

As one of the leading private research libraries in the United States, Northwestern University Library serves the educational and information needs of its students and faculty as well as scholars around the world. Visit the Library About page for more information or contact Distance Learning Librarian Tracy Coyne at 312-503-6617 or tracy-coyne@northwestern.edu.

Program-Specific Library Guides

- Business Databases A-Z
- Data Management
- Data Science
- Information Systems
- Sports Administration
- Statistics

Additional Library Resources

- Connectivity: Campus Wireless and Off-Campus Access to Electronic Resources
- Getting Available Items: Delivery to Long-Distance Patrons
- Quick Access to Major Newspapers
- Reserve a Library Study Room
- Resources for Data Analysis
- Schaffner Library Top Resources
- Sign up for an in-person or online Research Consultation Appointment
- Social Science Data Resources

The Writing Place

The Writing Place is Northwestern's center for peer writing consultations. Consultations are free and available to anyone in the Northwestern community: undergraduates, graduate students, faculty, or staff. To book an appointment, go to The Writing Place website.

The Math Place

The Math Place is a free tutorial service provided to students currently enrolled in Northwestern University's School of Professional Studies courses or in other Northwestern University courses. Students of all levels can benefit from the individual tutoring provided from this service, whether they are taking undergraduate or graduate level courses. To book an appointment, go to The Math Place website.

SPS Learning Studios

Learning studios are available to students who would like additional support in commonly used tools and topics, including: statistics, Excel, and coding in R. An instructor is available to answer your questions as you work through self-paced content and exercises. Students can self-enroll for free by visiting the SPS Academic Services page.

Read&Write Gold

Read&Write Gold is an optional text reading and writing program with numerous beneficial features. Originally developed to assist users with print disabilities, such as visual impairments, dyslexia, ADHD, etc., this program provides a wide array of tools to assist with reading, writing, and notetaking. One of the most useful tools is the text-to-speech function, which students may use to convert digital text into an audio format.

Read&Write Gold is available for free to all Northwestern students, faculty, and staff. Visit the Northwestern IT site on Read&Write Gold for more information about the software, as well as instructions on how to download it.

Academic Integrity at Northwestern

Students are required to comply with University regulations regarding academic integrity. If you are in doubt about what constitutes academic dishonesty, speak with your instructor or graduate coordinator before the assignment is due and/or examine the University website. Academic dishonesty includes, but is not limited to, cheating on an exam, obtaining an unfair advantage, and plagiarism (e.g., using material from readings without citing or copying another student's paper). Failure to maintain academic integrity will result in a grade sanction, possibly as severe as failing and being required to retake the course, and could lead to a suspension or expulsion from the program. Further penalties may apply. For more information, visit The Office of the Provost's Academic Integrity page.

Some assignments in SPS courses may be required to be submitted through Turnitin, a plagiarism detection and education tool. You can find an explanation of the tool here.

Course Technology

This course will involve a number of different types of interactions. These interactions will take place primarily through the Canvas system. Please take the time to navigate through the course and become familiar with the course syllabus, structure, and content and review the list of resources below.

Systems Requirements for Distance Learning

Students and faculty enrolled in SPS online classes should have access to a computer with the Minimum System Requirements.

Canvas

The Canvas Student Center includes information on communicating in Canvas, navigating a Canvas course, grades, additional help, and more. The Canvas at Northwestern website provides information of getting to know Canvas at Northwestern and getting Canvas support. The Canvas Student Guide provides tutorials on all the features of Canvas. For additional Canvas help and support, you can always click the Help icon in the lower left corner to begin a live chat with Canvas support or contact the Canvas Support Hotline.

The Canvas Accessibility Statement and Canvas Privacy Policy are also available.

Zoom

We will use Zoom for optional synchronous meetings. The Zoom support page provides additional guidance for using Zoom.

The Zoom Privacy Policy and the Accessibility Features on Zoom are also available.

Please note that any scheduled synchronous meetings are optional. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation. These synchronous sessions will be recorded, so you will be able to review the session afterwards.

LinkedIn Learning

Some required resources used in this course will be accessed via LinkedIn Learning. Northwestern offers access to LinkedIn Learning courses for students, faculty, and staff. Courses cover a wide range of technology and business topics including databases, programming, and data analysis. For more information, see the Human Resources LinkedIn Learning page.

O'Reilly Books Online

Some required readings and resources used in this course will be accessed via O'Reilly Books Online. Northwestern University Libraries provides access to O'Reilly Books Online. This includes many popular books on programming, software development, and other technology topics. It also includes online tutorials. Students can register for an account with a Northwestern email address; access the link from the Northwestern network or VPN.

Python

This course will use the Python programming language. Python can be downloaded at Anaconda. Download and install the Python 3.x version of Anaconda. Additional documentation on downloading and installing Anaconda can be found at Anaconda Documentation. After installation, you should be able to access the Anaconda Navigator. This course will be using Jupyter notebooks to build and store Python files for assignments. Jupyter can be launched via the Anaconda Navigator.

Required Technical Skills

Students in an online program should be able to do the following:

- Communicate via email and Canvas discussion forums.
- Convert Jupyter notebook files to HTML files.
- Use web browsers and navigate the World Wide Web.
- Use the learning management system Canvas.
- Use integrated Canvas tools.

Required Digital Literacy Skills

In order to be successful in an online course, students should be able to locate, evaluate, apply, create, and communicate information using technology.

Students in this online course should be able to do the following:

- Create, name, compose, upload, and attach documents.
- Download, modify, upload, attach document templates.
- Create, name, design, and upload files.
- Access and use a digital textbook.
- Search the Internet strategically and assess the credibility of Internet sources.
- Participate in threaded discussions by contributing text responses, uploading images, sharing links.
- Coordinate remote work with peers, which may include contacting each other by e-mail, phone, video conference, or shared document.

Technical Help and Support

The SPS Help Desk is available for Faculty, Students and Staff to support their daily IT needs. For additional technical support, contact the Northwestern IT Support Center.

Permissions

Instructional Materials

This course was developed in partnership with Distance Learning staff in the School of Professional Studies at Northwestern University. Every effort has been made to responsibly acquire instructional materials for this class, by adhering to copyright law, obtaining permission from copyright holders, selecting Open Educational Resources (OERs) and Creative Commons (CC) materials, and using citations to credit the work of others.

The same is expected of students in this course. Please review the Academic Integrity statement for more information.

Sharing Course Content

Content within this course--including assignment descriptions, exam questions, and other course components--may not be distributed outside of the course, either to other students or on the Internet more broadly.

Student Ownership of Content

Students retain ownership of all content developed while completing this course, as dictated by the university Copyright Policy ("copyright ownership resides with the Creator(s) of copyrightable works").

Per the Family Educational Rights and Privacy Act (FERPA), if your instructor wishes to share your work with future students, your permission must be obtained in writing.

Your instructor may limit access to the course after a cutoff date. When you complete the course, please ensure that you have saved all work. You may not be able to return to the course to download your submissions.

Course Schedule

Module 1

Learning Objectives

- Install and explore Anaconda Distribution
- Explore Jupyter Notebook and the command prompt
- Practice basic print commands and debugging techniques
- Describe the overall structure of Python and its benefits

Readings & Media

Required Readings

- Handout - *Getting Started in MSDS 430*
- Handout – *Jupyter Notebook Introduction*
- How to Think Like a Computer Scientist: Interactive Edition – Chapter 1

Required Media

- LinkedIn Learning – *Python Essential Training*. Listen to the following videos:
    - Introduction
    - Chapter 2 – Language Overview

Optional Media

- LinkedIn Learning – *Learning Python*. Consider listening to the following videos:
    - Introduction
    - Getting Started
    - Building Hello World

- O'Reilly Books Online – Jupyter Notebook Playlist

Assignments

- Introductions discussion (not graded)
- Module 1 Discussion
- Module 1 Python Assignment

Module 2

Learning Objectives

- Determine and recognize data types
- Write and execute conditional statements
- Perform mathematical calculations using operators
- Discuss the algorithmic process of programs

Readings & Media

Required Readings

- How to Think Like a Computer Scientist: Interactive Edition – Chapters 2 and 7

Optional Readings

- The Python Tutorial – Chapters 3 and 4 (4.1 – 4.5)
- How to Think Like a Computer Scientist: Interactive Edition – Chapter 3

Required Media

- LinkedIn Learning – *Python Essential Training*. Listen to the following videos:
    - Chapter 3 – Types and Values
    - Chapter 4 – Conditionals
    - Chapter 5 – Operators

Optional Media

- LinkedIn Learning – *Learning Python*. Consider listening to the following videos:
    - Variables and Expressions
    - Conditional Structures

Assignments

- Module 2 Discussion
- Module 2 Python Assignment

Module 3

Learning Objectives

- Create and execute loops
- Perform calculations using built-in functions
- Define new functions
- Investigate the importance of comments within a program

Readings & Media

Required Readings

- How to Think Like a Computer Scientist: Interactive Edition – Chapters 5, 6, and 8 (8.1 – 8.9)

Optional Readings

- The Python Tutorial – Chapters 4 (4.6 and 4.7), 10, and 11

Required Media

- LinkedIn Learning – *Python Essential Training*. Listen to the following videos:
    - Chapter 6 – Loops
    - Chapter 7 – Functions
    - Chapter 13 – Built-in Functions

Optional Media

- LinkedIn Learning – *Learning Python*. Consider listening to the following videos:
    - Python Functions

        - Loops

Assignments

- Module 3 Discussion
- Module 3 Python Assignment

Module 4

Learning Objectives

- Create and manipulate lists
- Read data from various file types
- Append data to files
- Write data to an output file
- Discuss the flexibility of reading files with Python

Readings & Media

Required Readings

- How to Think Like a Computer Scientist: Interactive Edition – Chapters 10 and 11

Optional Readings

- Reading Files with Python
- The Python Tutorial – Chapter 6

Required Media

- LinkedIn Learning – *Python Essential Training*. Listen to the following videos:
    - Chapter 12 – File I/O
    - Chapter 14 – Modules

Reference Material

- Pandas

Optional Media

- LinkedIn Learning – *Learning Python*. Consider listening to the following videos:
    - Chapter 4 – Working with Files

Assignments

- Module 4 Discussion
- Module 4 Python Assignment
- Project Proposal

Module 5

Learning Objectives

- Apply EDA techniques for cleaning data.
- Perform methods on data to determine its shape and structure.
- Analyze data to locate anomalies.
- Manipulate data to remove null, missing, and irrelevant values.

Readings & Media

Required Readings

- *The Data Science Method*
    - Problem Identification
    - Data Wrangling

Required Media

- O'Reilly Online Books – *Become a Python Data Analyst.* Listen to the following videos:
    - Chapter 3 – Pandas: Everyone's Favorite Data Analysis Library

Reference Material

- Numpy
- Pandas
- Python Exploratory Data Analysis Tutorial

Assignments

- Module 5 Discussion
- Module 5 Python Assignment

Module 6

Learning Objectives

- Apply visualization methods to recognize the distribution of data
- Examine relationships between variables
- Manipulate data as needed for EDA purposes
- Investigate tools for visualization

Readings & Media

Required Readings

- *The Data Science Method*

        - Exploratory Data Analysis

Required Media

- O'Reilly Online Books – *Become a Python Data Analyst*. Listen to the following videos:
    - Chapter 4 – Visualization and Data Analysis

Reference Material

- Matplotlib
- Seaborn
- Plotly

Assignments

- Module 6 Discussion
- Module 6 Python Assignment
- EDA Phase 1

Module 7

Learning Objectives

- Execute statements related to variables, expressions, and control flow
- Create formatted output and write to a file
- Manipulate text strings, convert input data, and perform calculations
- Investigate uses for extracting internet data

Readings & Media

Required Readings

- How to Think Like a Computer Scientist: Interactive Edition – Chapters 9 and 12

Required Media

- O'Reilly Books Online – *Python Programming Language*. Listen to the following videos:
    - Introduction
    - Lesson 1 – Working Environment
    - Lesson 2 – Program Structure and Execution Model
    - Lesson 3 – Text Processing and Files

Optional Media

- LinkedIn Learning – *Learning Python*. Consider listening to the following video:
    - Chapter 5 – Working with Web Data

Assignments

- Module 7 Discussion
- Module 7 Python Assignment

Module 8

Learning Objectives

- Investigate methods for organizing code
- Define functions and handle exceptions
- Build data structures from files
- Manipulate data from files

Readings & Media

Required Readings

- How to Think Like a Computer Scientist: Interactive Edition – Chapter 13 and Appendix: *Debugging*

Required Media

- O'Reilly Books Online – *Python Programming Language*. Listen to the following videos:
    - Lesson 4 – Functions and Error Handling
    - Lesson 5 – Data Structures and Data Manipulation

Optional Media

- LinkedIn Learning – *Python Essential Training*. Consider listening to the following video:
    - Chapter 10 – Exceptions

Assignments

- Module 8 Discussion
- Module 8 Python Assignment
- EDA Phase 2

Module 9

Learning Objectives

- Describe object-oriented programming
- Organize code bases into packages
- Create classes and attach methods
- Define and manipulate objects

Readings & Media

Required Readings

- How to Think Like a Computer Scientist: Interactive Edition – Chapters 17 & 18

Required Media

- O'Reilly Books Online – *Python Programming Language*. Listen to the following videos:
    - Lesson 6 – Library Functions and Import
    - Lesson 7 – Classes and Objects

Optional Media

- LinkedIn Learning – *Python Essential Training*. Consider listening to the following video:
    - Chapter 9 – Classes

Assignments

- Module 9 Discussion
- Module 9 Python Assignment

Module 10

Learning Objectives

- None

Assignments

- Module 10 Discussion
- EDA Report