

Syllabus & Course Schedule for DAAN 862



PennState
World Campus



DAAN 862: ANALYTICS PROGRAMMING IN PYTHON

COURSE DESCRIPTION

(3 credits) This course will explore the development of analytics systems and the application of best practices and established software design principles using the Python programming language and its several toolkits. Students will manipulate, analyze, and visualize complex data sets and implement statistical, machine learning, information visualization, text analysis, and social network analysis techniques through popular Python toolkits to gain insight into their data.

FACULTY



ENGINEERING DIVISION FACULTY, ADJUNCT INSTRUCTOR:
Prof. Guang Yang (*She/Her*)

Email: gfy5060@psu.edu (<mailto:gfy5060@psu.edu>)

OVERVIEW

This course will employ the Python programming language and its several toolkits for manipulating, analyzing, and visualizing complex data sets. Students will apply statistical, machine learning, information visualization, and text analysis techniques through popular Python toolkits such as Pandas, Matplotlib, Scikit-Learn, and NLTK to gain insight into their data.

OBJECTIVES

The objective of this course is to provide hands-on experience with data reading, cleaning, processing, and visualization techniques using Python programming language and its toolkits.



After successfully completing this course, students will be able to demonstrate:

- Understanding of the Python programming language and its development environment.
- The ability to manipulate, analyze, and visualize complex data sets using the Python language.
- Knowledge of the best practices and principles of software design in analytics systems.

COURSE MATERIALS



Required Textbook

- Book 1: ISBN: 978-1491957660 Wes McKinney, [Python for Data Analysis](https://wesmckinney.com/book/)  (<https://wesmckinney.com/book/>), O'Reilly, 3rd edition, 2017, the third edition is also available open-source at wesmckinney.com/book. It has an additional April 2023 update that coincides with the April 2023 release of Pandas 2.0.0, which includes new preprocessing functions like melt, pivot, pivot_table, and crosstab. If you are still using the 2nd edition, please note that the chapters after Chapter 12 are numbered differently in the 3rd edition.
- Book 2: ISBN: 978-1788299879 Gavin Hackeling, *Mastering Machine Learning with scikit-learn*, 2nd edition, 2017.
- Book 3: ISBN: 978-1484243534 Dipanjan Sarkar, *Text Analytics with Python*, 2nd edition, May 2019: <https://learning.oreilly.com/playlists/0e39264a-c60a-4732-ba91-a3cbfb602d85/>  (<https://learning.oreilly.com/playlists/0e39264a-c60a-4732-ba91-a3cbfb602d85/>).

Free Online Access to book chapters:

- You can buy hard copies of these books, or you can simply read them online by signing in to **O'Reilly Online Learning** (<https://www.oreilly.com/online-learning/>) with your Penn State email (i.e., xyz@psu.edu (<mailto:xyz@psu.edu>)). Click on **Sign In With Single Sign On** (SSO). Validate your authentication with WebAccess/ 2 Factors Authentication (DUO). Search the book by title or by providing the ISBN. Create a favorite list and add the book for quick access.
- You may purchase course materials from Barnes & Noble College (the bookstore used by Penn State's World Campus). For pricing and ordering information, please see the **Barnes & Noble College website** (<https://bncvirtual.com/psude>). Materials will be available at Barnes & Noble College approximately three weeks before the course begins. Alternatively, you may obtain these texts from other favorite bookstores. Be sure you purchase the edition/publication date listed.




Open Source Software

Anaconda Spyder: <https://www.anaconda.com/download/> 
(<https://www.anaconda.com/download/>)

TECHNICAL SUPPORT

Please remember that ALL questions about grades, course lesson content, and assignments should be directed to your course instructor.

If you have any technical difficulties using the tools within this course, please contact the Penn State Helpdesk.

- **HelpDesk Website:** <https://student.worldcampus.psu.edu/help-and-support/technical-support>  (<https://student.worldcampus.psu.edu/help-and-support/technical-support>)
- **HelpDesk Email:** techsupport@worldcampus.psu.edu
(<mailto:techsupport@worldcampus.psu.edu?subject=World%20Campus%20Tech%20Support%20Inquiry>)
- **HelpDesk Phone:** (800) 252-3592

CONTACT INFORMATION

All course-related e-mails should go through Canvas's course mail function (Canvas Inbox). Using Canvas to contact your instructor ensures that your message will be read, and your instructor will respond to you in a timely manner.

USING THE LIBRARY

Many of the University Libraries' resources can be utilized from a distance. Through the Library website, you can access magazines, journals, and articles; borrow materials and have them delivered to your doorstep; and get research help via email chat or phone from a librarian.

For more information, view the [Penn State University Library](https://libraries.psu.edu/)  (<https://libraries.psu.edu/>) website.

COURSE REQUIREMENTS AND GRADING

Students will be evaluated on their understanding of the course material by completing eleven (11) weekly assignments that evaluate their understanding of and ability to apply material contained in the lectures and reading assignments. One cumulative final exam will evaluate their knowledge of the Python language and program design principles, *Final Grades* will be calculated as follows:

Assignment Details

Assignment Category	Quantity	Points	Weight (% of Final Grade)
Lesson Assignments	11	100 (each)	75%
Project-Based Final Exam	1	100	25%

***Grades** will be based on the following scale:

A = 96 – 100, A- = 90 – 95, B+ = 87 – 89, B = 84 – 86, B- = 80 – 83, C+ = 77 – 79, C = 70 – 76, D = 60 – 69, and F = 60 and below.



Homework Assignments and Project-Based Final Exam

There will be 11 assignments and 1 project-based final exam.

- All assignments and the project-based final project should be solved by Python instead of other software.
- Please submit a Word or PDF file for all assignments and the project-based final project.
- All assignments and the project-based final project should be completed independently without collaboration with other students.
- You will have a week to finish each assignment, and they are due at 11:59 PM Eastern time on the date specified.
- Late submissions will not be accepted unless the instructor has given permission prior to the due date.

After you have submitted an assignment, you will typically be able to review your grade and any comments made by your instructor within 7 days after the due date. This process is used for all homework assignments, exams, or other graded submissions. Some instructors may also send you a message informing you that the assignment has been graded. Some instructors may choose to release all the grades to all students at once; other instructors may release grades per student one at a time.

UNIVERSITY POLICIES AND RESOURCES

Be sure to review the [University Policies and Resources](https://sites.psu.edu/gvwc/course/penn-state-university-policies-and-resources/) 

(<https://sites.psu.edu/gvwc/course/penn-state-university-policies-and-resources/>), which include important information regarding academic integrity, student disability resources, educational equity, counseling services, and technical requirements.

COURSE SCHEDULE

The schedule below outlines the topics we will be covering in this course, along with the associated time frames and assignments.

Please note: All submissions must be submitted to the proper assignment in Canvas prior to 11:59 PM EST on the final day of the lesson week.

- **Course begins:** August 26, 2024

- **Course ends:** December 13, 2024

Lesson 1: Introduction to Python

<i>Timeframe:</i>	August 26 - September 1
<i>Readings:</i>	Textbook <ul style="list-style-type: none">◦ Python for Data Analysis, Wes Makinney, 2nd edition.<ul style="list-style-type: none">▪ Chapter 2▪ Chapter 3
<i>Activities and Assignments:</i>	<ul style="list-style-type: none">• Lesson 1 Assignment<ul style="list-style-type: none">• Assignment due by 11:59pm EST Sunday, September 1

Lesson 2: Numeric Analysis with Numpy

<i>Timeframe:</i>	September 2 - September 8
<i>Readings:</i>	Textbook <ul style="list-style-type: none">• Python for Data Analysis, Wes Makinney, 2nd edition.<ul style="list-style-type: none">◦ Chapter 4
<i>Activities and Assignments:</i>	<ul style="list-style-type: none">• Lesson 2 Assignment<ul style="list-style-type: none">• Assignment due by 11:59pm EST Sunday, September 8

Lesson 3: Statistical Analysis with Pandas

<i>Timeframe:</i>	September 9 - September 15
<i>Readings:</i>	Textbook <ul style="list-style-type: none"> • Python for Data Analysis, Wes Makinney, 2nd edition. <ul style="list-style-type: none"> ◦ Chapter 5 ◦ Chapter 6
<i>Activities and Assignments:</i>	<ul style="list-style-type: none"> • Lesson 3 Assignment <ul style="list-style-type: none"> • Assignment due by 11:59pm EST Sunday, September 15

Lesson 4: Data Cleaning, Processing and Manipulating with Pandas

<i>Timeframe:</i>	September 16 - September 22
<i>Readings:</i>	Textbook <ul style="list-style-type: none"> ◦ Python for Data Analysis, Wes Makinney, 2nd edition. <ul style="list-style-type: none"> ▪ Chapter 7
<i>Activities and Assignments:</i>	<ul style="list-style-type: none"> • Lesson 4 Assignment <ul style="list-style-type: none"> • Assignment due by 11:59pm EST Sunday, September 22

Lesson 5: Data Wrangling with Pandas

<i>Timeframe:</i>	September 23 - September 29
--------------------------	-----------------------------

Readings:	Textbook <ul style="list-style-type: none"> Python for Data Analysis, Wes Makinney, 2nd edition. <ul style="list-style-type: none"> Chapter 8
Activities and Assignments:	<ul style="list-style-type: none"> Lesson 5 Assignment <ul style="list-style-type: none"> Assignment due by 11:59pm EST Sunday, September 29

Lesson 6: Data Visualization with Matplotlib

Timeframe:	September 30 - October 6
Readings:	Textbook <ul style="list-style-type: none"> Python for Data Analysis, Wes Makinney, 2nd edition. <ul style="list-style-type: none"> Chapter 9
Activities and Assignments:	<ul style="list-style-type: none"> Lesson 6 Assignment <ul style="list-style-type: none"> Assignment due by 11:59pm EST Sunday, October 6

Lesson 7: Evaluation Models with Scikit-Learn

Timeframe:	October 7 - October 13
Readings:	Textbook <ul style="list-style-type: none"> Mastering Machine Learning with Scikit-Learn, 2nd edition, Gavin Hackeling

	<ul style="list-style-type: none"> ▪ Chapter 1
Activities and Assignments:	<ul style="list-style-type: none"> • <i>No Assignment</i>

Lesson 8: Supervised Learning with SciKit-Learn I: Classification Models

Timeframe:	October 14 - October 20
Readings:	Textbook <ul style="list-style-type: none"> • Mastering Machine Learning with Scikit-Learn, 2nd edition, Gavin Hackeling <ul style="list-style-type: none"> ◦ Chapter 6 ◦ Chapter 7 ◦ Chapter 8 ◦ Chapter 10
Activities and Assignments:	<ul style="list-style-type: none"> • Lesson 8 Assignment <ul style="list-style-type: none"> • Assignment due by 11:59pm EST Sunday, October 20

Lesson 9: Supervised Learning with SciKit-Learn II: Regression

Timeframe:	October 21 - October 27
Readings:	Textbook <ul style="list-style-type: none"> • Mastering Machine Learning with Scikit-Learn, 2nd edition, Gavin Hackeling <ul style="list-style-type: none"> ◦ Chapter 6 ◦ Chapter 7

	<ul style="list-style-type: none"> ◦ Chapter 8 ◦ Chapter 10
Activities and Assignments:	<ul style="list-style-type: none"> • Lesson 9 Assignment <ul style="list-style-type: none"> • Assignment due by 11:59pm EST Sunday, October 27

Lesson 10: Supervised Learning with SciKit-Learn III: Advanced Models

Timeframe:	October 28 - November 3
Readings:	Textbook <ul style="list-style-type: none"> • Mastering Machine Learning with Scikit-Learn, 2nd edition, Gavin Hackeling <ul style="list-style-type: none"> ◦ Chapter 9 ◦ Chapter 11
Activities and Assignments:	<ul style="list-style-type: none"> • Lesson 10 Assignment <ul style="list-style-type: none"> • Assignment due by 11:59pm EST Sunday, November 3

Lesson 11: Unsupervised learning with Scikit-Learn: Clustering


Timeframe:	November 4 - November 10
Readings:	Textbook <ul style="list-style-type: none"> • Mastering Machine Learning with Scikit-Learn, 2nd edition, Gavin Hackeling <ul style="list-style-type: none"> ◦ Chapter 13

Activities and Assignments:	<ul style="list-style-type: none"> • Lesson 11 Assignment <ul style="list-style-type: none"> • Assignment due by 11:59pm EST Sunday, November 10
------------------------------------	--

Lesson 12: Text Mining with re and NLTK

Timeframe:	November 11 - November 17
Readings:	Textbook <ul style="list-style-type: none"> • Text Analytics with Python, Dipanjan Sarkar, 2016 <ul style="list-style-type: none"> ◦ Chapter 1 ◦ Chapter 3
Activities and Assignments:	<ul style="list-style-type: none"> • Lesson 12 Assignment <ul style="list-style-type: none"> • Assignment due by 11:59pm EST Sunday, November 17

Lesson 13: Classification of Text

Timeframe:	November 18 - November 24
Readings:	Textbook <ul style="list-style-type: none"> • Text Analytics with Python, Dipanjan Sarkar, 2016 <ul style="list-style-type: none"> ◦ Chapter 5
Activities and Assignments:	<p>*Prepare for the Final Exam.</p> <hr/> <p>***Please also remember to complete the online SEEQ evaluation  (https://rateteaching.psu.edu/) for</p>

	<i>your course instructor.</i>
--	---------------------------------------

*PSU Official Fall Break/Thanksgiving Holiday (NO CLASSES)

<i>Timeframe:</i>	November 24 - November 30
--------------------------	---------------------------

Lesson 14: Final Project

<i>Timeframe:</i>	December 2 - December 13
<i>Readings:</i>	(None)
<i>Activities and Assignments:</i>	<ul style="list-style-type: none"> • Your project-based final exam is due by 11:59pm EST Sunday, December 8.

***Note:** *If you are planning to graduate this semester, please communicate your intent to graduate to your instructor. This will alert your instructor to the need to submit your final grade in time to meet the published graduation deadlines. For more information about graduation policies and deadlines, please go to the **Graduation Information** (<http://student.worldcampus.psu.edu/courses/graduation>) on the My Penn State Online Student Portal.*

*Formal instruction will end on the last day of class. Provided that you have an active Penn State Access Account user ID and password, you will continue to be able to access the course materials for one year, starting from the end date of the academic semester in which the course was offered (with the exception of library reserves and other external resources that may have a shorter archival period). After one year, you might be able to access the course based on the policies of the program or department offering the course material, up to a maximum of three years from the end date of the academic semester in which the course was offered. For more information, please review the **University Course Archival Policy** (<https://canvas.psu.edu/2019/08/20/canvas-archive-policy-for-lionpath-courses/>).

Disclaimer: Please note that the specifics of this Course Syllabus are subject to change, and you will be responsible for abiding by any such changes. Your instructor will notify you of any changes.

