BAIS:6400:0EXF | Spring 2021

Data Programming in Python

|  |  |
| --- | --- |
| Course Instructor David Thole  Campus Address: -  Phone: by-email  Email: david-thole@uiowa.edu  Office Hours: By appointment or via email | Class Meeting Times This course is organized within a structure of scheduled course work. Students will progress through the course as a cohort and will complete and/or submit course work online. Recorded lectures will be viewable at choice of time within week. Required (online) class sessions are a part of this course. |
| Academic Course Home [Tippie College of Business](https://tippie.uiowa.edu/)  DEO: JProf. Ann Campbell  Phone: (319) 335-0918  Email: [ann-campbell@uiowa.edu](mailto:ann-campbell@uiowa.edu) | Course Site  To access the course site, log into [Iowa Courses Online (ICON)](http://icon.uiowa.edu/index.shtml) using your Hawk ID and password. Prerequisites None |

Program Goals

The Tippie MSBA has learning goals that drive decisions about curriculum and assignments within courses:

Program Goal 1: Graduates will exhibit knowledge and skills relevant to data and its application in business

Program Goal 2: Graduates will create and communicate solutions to data-related business problems that impact their organizations and communities

Program Goal 3: Graduates will understand and contemplate ethical and privacy issues arising in their own work

Program Goal 4: Graduates will demonstrate the ability to be effective team members in a diverse and complex world.

The primary emphasis on this course is goal #1 and #4. Assignments will challenge you to develop skills in Python (goal #1). The class project will allow for a more ethical understanding of complex analysis (goal #1, #2, #3), and will encourage working effectively in groups (goal #4)

Course Description and Goals

This course introduces the principles and practices of handling, cleaning, processing, modeling, and visualizing data using the Python programming language within the context of building a working machine learning application in Flask. The course places a special emphasis on techniques and tools useful for data scientists, specifically the data exploration and cleaning aspects. Accordingly, the main programming environment for the course will be the open-source Python environment using Jupyter Notebook. You may also find it useful to use an integrated development environment (IDE) such as Spyder, VS Code, or PyCharm for application development. However, an IDE is not required. The main topics include:

* Introduction to Python and Jupyter Notebook/Hub
* Python Basics
* Data Manipulation and Analysis
* Files and External Data Sources
* Data Visualization
* Machine Learning and Python Flask Application Development
* Time Series Analysis
* Analytics in the Cloud (Azure, AWS) and Docker

Media/System Requirements

Technical requirements for completing this online course include:

* Student-provided personal computer. Any operating system is fine, and support can be given regardless of the device. Instruction, in videos, will be given in Windows. Please avoid basing all your work on something like an iPad or Android tablet/phone. A laptop, desktop running Windows, Linux (x64), or OSX is your best bet.
* Computer with reliable Internet access. A wired Ethernet connection to the internet is very strongly suggested. Wireless and cellphone data connections may experience connection problems. Android and iOS operating systems are not fully supported at this time. See specific requirements on the [Distance and Online Education Technical Requirements/Download page](https://distance.uiowa.edu/courses/technical-support).
* USB/Bluetooth headset with microphone and a web camera capable of conferencing. Most laptops have a built in camera which typically works fine. However, a USB headset with microphone is recommended to minimize audio feedback and room noise and maximize sound quality.
* While tablets, smartphones and other mobile devices may allow for some completion of coursework, they are not guaranteed to work in all areas. Please ensure you have a Windows, Mac, or Linux based computer available to complete coursework in the event your selected mobile device does not meet the needs of the course.

Students who need assistive technologies will have different computer and technology requirements. Please check with your [Student Disability Services](https://sds.studentlife.uiowa.edu/) to determine the requirements for the specific technologies needed to support your online classes.

For questions regarding virtual classrooms (i.e. Zoom) or UICapture (Panopto) please contact [Continuing Education Technical Support](mailto:dce-techsupport@uiowa.edu) (319-335-3925).

Need help with ICON or your HawkId? Please contact the [ITS Helpdesk](mailto:ITS%20Help%20Desk%20%3cits-helpdesk@uiowa.edu%3e) (319-384-HELP).

Required Textbook/Media

There are no required texts for the course. However, the following books are incredibly useful, in order of usefulness, and will be referenced:

**The Quick Python Book**

**(**[**https://www.manning.com/books/the-quick-python-book-third-edition**](https://www.manning.com/books/the-quick-python-book-third-edition)**)**

Python for Data Analysis by Wes McKinney

(<https://www.oreilly.com/library/view/python-for-data/9781491957653/>)

The Internet is also a rich resource for the course. In addition, the course web site will be hosted on the university ICON server (listed above). Please check the web site frequently for announcements, assignments, etc.

Grading Criteria

The coursework breakdown is as follows:

* 40% graded homework assignments
* 30% exams (two programming exams, see calendar)
* 25% group project
* 5% group participation rating

Final grades will be assigned according to the following planned curve, which has been approved by the Business Analytics Graduate Program (however, the instructor reserves the right to deviate from this curve if deemed necessary):

* A range: ≈ 50% of students
* B range: ≈ 50% of students
* C, D, F: as needed

In addition, the A range and B range may be equally divided into +/- designations. While an A+ is possible, it will be reserved only for times when clear mastery has been exhibited across all topics, and when the grade ranks allow for a clear break between an A and A+.

Course Structure

The instructor will provide recorded lectures and practice exercises on ICON for each module. There will be a 90-minute live Zoom session each week on the assigned evening. The purpose of the live sessions is to review key points from the recorded lectures and answer student questions regarding the lectures, practice exercises, homework, or group project. Time permitting, the live sessions will include breakout sessions to work on the group project.

This course is being offered over the World Wide Web. Students will login to the course site on ICON to access the course materials. For details of the course assignments and activities, see the “Course Work” section of this syllabus.

Students are expected to visit the course site regularly to:

* Access assigned course materials (posted on the “Modules” page) such as pre-recorded lectures and journal articles.
* Review the course homepage regularly for any updates related to the course “Announcements” and/or “Calendar.”
* Submit assignments to the course instructor via the ICON “Assignments.”
* Participate in the “Discussion” forums.

Course Work

Online Discussions/Participation (5%):

The lectures are designed for more discussion and clarifying. Participation in this regard means asking questions, providing ideas, and so on. Furthermore, the forum is a great place when people have questions. Participate, help your fellow student, etc.

Application Assignments (8x5%):

There are 8 total programming assignments. These will be posted a week before they are due. While they’re not intended to be very difficult, I encourage exploration and going “beyond” what’s directly requested. The idea is to use the homework assignment as a method to explore what was discussed.

Homework will be returned within a week after the due date. This will be provided in the same Jupyter notebook with markdown-based comments/annotation.

Exams (2 x 15%):

There are two programming-based exams in the course. These will be provided a week before they are due. These will be more difficult than the homework, but still respectful of time constraints. Given the online nature of this course, it’s recommended you limit use of answers from sites such as StackOverflow. Grading on the exams will take in account creativity, and a demonstration of your true knowledge.

Group Project

The goal of the course is to learn the python language in the context of creating an operational predictive model within a group setting. This allows students the opportunity to apply data science techniques using python. The driver for this learning environment is a course-long group project that incorporates gathering, cleaning, transforming, and exploring data; followed by fitting, tuning, and building a model that is operationalized in a python Flask application.

The project includes creating one or more Jupyter notebooks that encompasses testing, data gathering, cleaning, transforming, and exploring the data; and another notebook that encompasses model creation, summary and so on. It’s optional to also create libraries that the notebooks can consume (thus greatly cleaning up the notebooks). There will be a checkpoint mid-semester to review the group’s progress to date.

Each group will submit a working application file, a test observation for the application, the input data for the project, and the project notebooks all in a zip file. The expectation is a working Flask application that at a minimum returns JSON.

Course-Specific Policies and Guidelines

As a registered student you are responsible for the course policies posted below.

**Communications:** Students can expect to receive weekly communications from the instructor (via course “Announcements”). Students are also responsible for all official correspondence sent through their University of Iowa email address. Privacy considerations, such as federal law, may apply when using an address other than the standard University e-mail address. You can expect to receive responses to your inquiries within 24-48 hours. Please note that very rarely will emails be sent between 8AM-5PM M-F. Please do NOT wait til the last minute to contact on help needed. Often times emails will be replied to very early (between 3-6AM).

Office hours are not set, but time can be arranged. Most likely in-person meeting will be later in the week (Friday-Sunday).

**Exams:** Students are expected to take the exams at the regularly scheduled times, unless permission has been granted by the instructor at least 2 weeks in advance in writing. All exams are INDIVIDUAL WORK and collaboration on exams is not permitted under any circumstance. All exams are CLOSED BOOK. Use of any materials during the exam is strictly prohibited and will be grounds for a grade of zero on the exam.

**Attendance**: Attendance in all class sessions is expected. However, unforeseen conflicts due occasionally arise. Students are expected to give adequate advance notice of an excused absence, make up the missed work and make every effort to avoid a class conflict. To ensure legitimacy, students may be asked to provide documentation upon request. Short-term illnesses will also be considered as excused absences, however students are asked to let me know prior to or even the day of the absence why they are not in class. A plan for managing longer term absences should be discussed with me to determine whether successful completion of the course is possible.

**Late Work**: Because I regard this class as I would any job responsibility, I hold students to strict standards of timeliness and ethics. In this regard, late work is unprofessional and will be penalized. Assignments that are late will receive an automatic 20% grade reduction. Each following day late will result in another 20% reduction in the grade.

**Collaboration**: I encourage discussion and collaboration, but all written assignments must be written by the student (or students in the case of group project) alone. I encourage students to get feedback from others about papers but no one should be writing full sentences or paragraphs except for the student(s) whose name(s) appears on the paper. I expect that you will refrain from plagiarizing others’ work, either in this class or in other classes. I will be using Turnitin, which is a plagiarism detection program. During exams, no outside notes or materials may be used.

**Assignment Format:** Please submit all your homework in Jupyter notebook format, as much as possible. There may be some cases where a Jupyter notebook is not appropriate, in which case, please provide a zip file of your source. In those cases, please add a requirements.txt file that lists the libraries you’re using.

**Netiquette:** The term “netiquette” refers to the do’s and don’ts of online communication. As it applies to this online course, it is my expectation that students will communicate effectively and respectfully with each other, the instructor, and our guest speakers (if applicable). [Follow this link to learn more about The Core Rules of Netiquette](http://www.albion.com/netiquette/corerules.html). <http://www.albion.com/netiquette/corerules.html>

**Inclement Weather/Class Cancellation**: Although it is our intent to offer every class at its assigned time, on rare occasion there are weather or other emergency events that require that alternative arrangements are made for class delivery. Because these sessions are online, even if Tippie College or University in-person classes are cancelled, I anticipate holding the regularly scheduled Zoom session of the class. If a situation occurs such that Zoom cannot be used, I will cancel the class completely – all reading and work that was due is expected to be completed and turned in on time through ICON.

Collegiate and University Policies

The administrative home of this course is the Tippie College of Business, which governs academic matters relating to the course such as the add/drop deadlines, issues concerning academic misconduct, and how credits are applied for various graduation requirements. Different colleges might have different policies.

As a registered student in a course in the Tippie College of Business, you are responsible for the collegiate policies posted at <https://tippie.uiowa.edu/collegiate-policies-mba>.

As a registered student in a course through The University of Iowa, you are responsible for the University policies posted at <https://provost.uiowa.edu/sites/provost.uiowa.edu/files/wysiwyg_uploads/Syllabi%20updates%20for%20spring%202021.pdf>.

Course Calendar

This online course is organized within a structure of scheduled course work (see weekly calendar below). Although you are not required to meet in a classroom, we will progress through the course materials as a class. For this reason, you must manage your time effectively in order to the complete the assigned course work according to the firm due dates listed in the calendar below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Module** | **Lecture Date** | **Topic** | **Due** |
| 0 | 01/25 | Introduction to Python, Jupyter Notebook, and Anaconda  Project Announcement. |  |
| 1 | 01/25 | Python Basics Part 1: Data Types  (numbers, strings), basic operators |  |
| 2 | 02/01 | Python Basics Part 2: Advanced Data Types (lists, tuples, dictionaries, sets), Flow Control | HW 1 Due (02/01) |
| 3 | 02/08 | Functions, Modules, and Exceptions  And  Handling Numbers with Numpy | HW 2 Due (02/08) |
| 4 | 02/15 | Data Manipulation and Analysis with Pandas | HW 3 Due (02/15) |
| 5 | 02/22 | Files and External Data Sources | HW 4 Due (02/22) |
| 6 | 03/01 | Data Visualization with Matplotlib, Seaborn, and Plotly | Exam 1 available online – Due Mar 7 @ 11:59 PM |
| 7 | 03/08 | Advanced Data Analytics: Machine Learning with Scikit-Learn | Checkpoint – Jupyter Notebook due Mar 8.  HW 5 Due (03/08) |
| 8 | 03/15 | Machine Learning using Pipelines and Grid Search  Setting up and Building a python Flask application | HW 6 Due (03/15) |
| 9 | 03/22 | Time Series Analysis & Forecasting | HW 7 Due (03/22) |
| 10 | 03/29 | Introduction to computing in the Cloud (Azure) with Python, SQL Machine Learning Services. | HW 8 Due (03/29)  Teammate Feedback due April 2 |
| 11 | 04/05 | Introduction to computing in the Cloud (AWS) with Python, Lambda-like functions and Docker. | Project Due 04/05  Exam 2 available online – Due April 5 |