

## COSC 3360 – Intro to Data Science

**Instructor:** Dr. Mircea Agapie      **Office:** Hydrology 108      **Phone:** (254) 968-1820  
**E-mail:** [agapie@tarleton.edu](mailto:agapie@tarleton.edu)      Include course # in subject line, e.g. **3360 homework 1**  
**Website:** [www.agapie.net](http://www.agapie.net)      **Office hours:** Mon+Fri: 9-12, 2-5, Tue+Thu: 9-10, Wed: 2-5.

Meeting times and location: Tue and Thu 5-7:30 pm, in SCIEN 206

- We do not follow the usual lecture - lab format; instead each session alternates between lecture format (when we talk and solve pencil-and-paper quizzes) and lab format (when we code).
- We shall use Python exclusively as programming language, and we shall teach Python from scratch ... albeit in an accelerated fashion. This is not an intro to programming. If you have time and disposition, get a head start with any of the widely available online Python tutorials, first and foremost <https://docs.python.org/2.7/tutorial/index.html> - sections 3, 4 (skip 4.7), and, optionally 5. We shall use mostly Python 2 syntax, with the occasional pointer to Python 3.
- We shall use the Anaconda distribution, already installed in SCIENCE 206, with the Spyder editor. For practicing small Python snippets, you can also use any of the online interpreters listed on our webpage. It is recommended to install Anaconda on your home computer/laptop.

**Textbook:** There is no textbook required. I will distribute handouts at the beginning of each class and/or send them to you by email.

**Effort expected:** The average student will need **6 hours of individual work<sup>1</sup> per week** for this course. Please reserve these 6 hours on your weekly schedule. What you should do during this time:

- Immerse yourself in Python, code as much as you can.
- Study the material covered in class (take notes and use them!)
- Solve the individual work problems assigned for the next class, and homework problems.
- Redo all quiz problems.
- Develop your own memory-sheet for coding and algorithm topics.

**Keeping in touch:** the following channels of communication will be used. Please do your best to check them often (a few hours before class is a good idea):

- Email. Please check your Tarleton student email and make sure you clean your mailbox regularly – if it's full, the messages will bounce back.
- Messages posted on the course website [www.agapie.net](http://www.agapie.net)

### Grading:

- **Individual work and quizzes** (not graded): During most classes, short quizzes will be assigned to solve on the spot. At the end of class, the instructor will provide a list of problems which are recommended to solve before the next class. They are intended as a self-check for the student, to make sure the material covered was understood. Solve both quizzes and individual work in your notebook, or directly as Python programs.
- **Homework:** Six or seven sets of problems will be assigned over the entire semester, due in one week.
  - The homework is due at the beginning of class. Late submissions will not be credited. Exceptions are allowed only in extraordinary circumstances, if you have notified the instructor before the due date.
  - All homework will involve programming; give the instructor a hard-copy, but keep all your programs in electronic form for reference.
- **Exams:** Midterm: Tue, **Oct 16**, during the regular class. Final: Tue, **Dec 11** 6:30-9 PM.
- **Weights:** Homework 50%      Two exams 25% each
- **Intervals:** 90-100 = **A**, 80-90 = **B**, 70-80 = **C**, 60-70 = **D**, below 60 = **F**.

---

<sup>1</sup> Individual work does not include the lectures and labs.

- Up to 5% of the final score can be gained as extra-credit for answering questions in class or lab, or for solving homework problems assigned as extra-credit.

**Intended Student Learning Outcomes:** At the conclusion of the course the student will:

- Be able to define data science, and know the job responsibilities of a data scientist, as opposed to those of professionals in other data-related fields.
- Have a working knowledge of the various parts of the “data pipeline”.
- Be familiar with the history of the data science field, its typical applications, and the job prospects for data scientists.
- Have solid Python programming skills, including the fundamental Python data types (lists, strings, tuples, dictionaries, files), list comprehensions, lambda expressions, and regular expressions.
- Be able to download, open, and extract data from files and data streams of various popular formats (CSV, XLS, HTML, XML, JSON).
- Be able to acquire data using a web service (Google Geodata).
- Be able to perform text analysis using these tools:
  - Python’s file object and primitive string operations
  - The counter class from Python’s Collections module
  - Regular expressions from Python’s re module
- Be able to perform numerical analysis using these tools:
  - Python’s numpy module for matrix computations
  - Python’s pandas module for labeled data, time-series, statistics.
- Be able to visualize data with Python’s module matplotlib
- Have a general knowledge of the classes of machine learning algorithms. Be able to use in a program one machine learning algorithm in each of these classes:
  - Classification
  - Clustering
  - Regression

**Academic Conduct:** Students guilty of academic dishonesty, cheating, or plagiarism in academic work shall be subject to disciplinary action. The instructor may initiate disciplinary action in any case of academic misconduct.

- For homework and lab work, it is allowed (actually recommended) for students to brainstorm and develop solutions together, however, **the writing and coding have to be done individually.**
- **It is forbidden to copy solutions or code found online.**

**Absence Policy:** Attendance will not be recorded for this course. The student is responsible for any material covered in class and in the lab, and for obtaining assignments and other materials for classes from which they are absent.

**Make-up Policy:** The instructor has the responsibility and authority to determine if work can be made-up because of absences. Students may request to make-up work for valid and verifiable reasons such as: illness, death in the immediate family, legal proceedings, or participation in sponsored University activities. Let instructor know in advance of any absence/make-up issues. **No make-up requests after the due date!**

**Students with disabilities:** It is the policy of Tarleton State University to comply with the Americans with Disabilities Act and other applicable laws. If you are a student with a disability seeking accommodations for this course, please contact the Center for Access and Academic Testing at 254.968.9400 or caat@tarleton.edu. The office is located in Math 201. More information can be found at [www.tarleton.edu/caat](http://www.tarleton.edu/caat), in the University Catalog, or at [www.ada.gov](http://www.ada.gov).