Course Information

Course Number:	STAT 315/ECEN 360
Course Title:	Computational Data Science
Section:	500 (in-person, synchronous)
Time:	MWF 11:30am-12:20pm Central Time (CT)
Location:	Blocker 411
Credit Hours:	3.0

Instructor Details

Instructor:	Scott A. Bruce
Office:	Blocker 406D
Phone:	979-845-3141 (main office)
E-Mail:	sabruce@tamu.edu
Office Hours:	M 1:00pm-2:00pm CT online via Zoom or in-person by appointment; other times can be arranged upon request

Teaching Assistant Details

TA:	Ramana Math
E-Mail:	ramana_hm@tamu.edu
Office Hours:	T 2:00pm-3:00pm CT online via Zoom by appointment; other times and in-person meetings can be arranged upon request

Important Dates

Exam Dates and Times

Midterm Exam: Wednesday, March 6th 11:30am-12:20pm CT
Final Report and Presentations: TBD

University Dates and Times

Last day for adding/dropping courses: Monday, January 22nd 5:00pm CT
Spring break: Monday March 11th – Friday March 15th
Last day to drop courses with no penalty (Q-drop): Tuesday, April 16th 5:00pm CT
Last day to officially withdraw from the University: Tuesday, April 16th 5:00pm CT
Last day of class: Tuesday, April 30th (Spring Redefined Day)

Course Description

Data science is a multidisciplinary field that utilizes statistics, data analysis, machine learning, algorithms, software and computing systems to extract information, acquire knowledge, and gain insights into the underlying context from which data is generated. The course introduces the computational practice of data science through a sequence of interactive modules that provide an integrated hands-on approach to its methods, tools, and applications, and supporting technologies including high performance and cloud computing platforms.

The course is aimed both at students who wish to acquire knowledge of data science by developing fluency in its applications and typical workflows, and also students with previous exposure to data science foundations who wish to develop complementary skills in the use of state-of-the-art systems and tools. Python will be used, whenever possible, to interact with the tools covered.

Topics covered span the entirety of the data life cycle and include relational databases and SQL; NoSQL databases; Python libraries frequently used by data scientists for data visualization, exploratory data analysis, and data modeling using common statistical learning, machine learning, and deep learning techniques; Git, GitHub, and Docker for development and version control; using High Performance Computing (HPC) clusters; parallel distributed processing in Python via Dask (time permitting).

Course Prerequisites

1. Grade of C or better in ENGR 102 or prior programming experience
2. Grade of C or better in MATH 251, MATH 253, or STAT 211
3. Junior or senior classification

Course Learning Outcomes

- Understand fundamental concepts of relational databases (e.g. data types, keys, relations).
- Explain the difference between relational and non-relational databases and identify appropriate candidate database solutions for different use cases.
- Use SQL to summarize and identify patterns in data.
- Construct SQL queries from a given prompt and database.
- Use Python data science libraries to access and analyze data.
- Use Git and Github for collaborative code development and version control.
- Create and use Docker containers to facilitate sharing of data science projects.
- Use distributed data processing tools such as Dask to manipulate and analyze data that cannot be stored in memory all at once.
- Design non-trivial data science projects with Python comprising data cleaning and pre-processing, data summarization and visualization, and creating, training, and deploying common machine learning and deep learning models.

Textbook and Resource Materials

Required Textbooks

1. *SQL for Data Analytics* by Upom Malik, Matt Goldwasser, and Benjamin Johnston. Packt Publishing, 2019. ISBN: 978-1-78980-384-6
2. *The Python Workshop* by Andrew Bird, Lau Cher Han, Mario Corchero Jiménez, Graham Lee, and Corey Wade. Packt Publishing, 2019. ISBN: 978-1-83898-453-3
3. *An Introduction to Statistical Learning with Applications in Python* by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. Springer, 2023. ISBN: 78-3-031-38747-0

These textbooks are freely available from the library using the links above after logging in with your TAMU username and password.

Required Online Resources

DataCamp interactive video tutorials will be assigned throughout the course, and DataCamp workspaces will be used for collaborative code development. Please use this link to create a free account using your tamu.edu e-mail address and name as it appears in Canvas and register for this course to access these materials. Do not share this link with anyone outside the course.

LinkedIn Learning video tutorials may also be assigned throughout the course. You are able to access these materials freely by logging in using your TAMU username and password.

GitHub Classroom will be used for some group assignments to facilitate collaboration and track individual contributions towards group submissions. If you haven't do so already, please create a free GitHub account and then use this link to join the GitHub classroom for the course.

iClicker will be used to track attendance. You will need to download and install the iClicker app on your phone in order to register your attendance each class period. Use this link to join the course in iClicker via join code **WVBO**.

The online portion of this course will require a significant time commitment. Give yourself plenty of time to complete the online materials and related assignments.

Course Communication Policy

Canvas will be the primary source of information relevant to the course (e.g. announcements, handouts, assignments, changes to office hours). You are expected to check the course Canvas site at least three times a week and whenever an announcement is posted.

If you have a question that arises during the class, please ask in class! This helps the instructor regulate the pace of the course and address issues along the way. Time permitting, the question can be addressed immediately.

When seeking assistance with the course outside of class, please follow these steps.

1. If the question may be helpful for others in the class, and can be easily posed and answered online, post your question on the appropriate discussion board in Canvas. The TA, instructor, or other students can respond to your post in Canvas, so others with the same questions can also see the answers.
2. If the question may be helpful for others in the class, but you are having trouble communicating your question in writing, or if you need one-on-one guidance, drop in during office hours so the instructor or TA can work with you individually.
3. If 1-2 don't apply to your particular question, e-mail the TA for assistance. Please include 'STAT 315' at the beginning of the subject line. Please allow for at least 2 business days for a response.

Talking during class lectures is not permitted. Please raise your hand to ask or answer questions. If you must carry on a conversation during class, please be courteous and take it out of the classroom.

Anonymous Feedback Tool

In order to improve students' educational experiences during the course, a completely anonymous feedback tool is provided to allow students to provide safe, actionable feedback starting from the first day of class. Some use case examples include:

- Teaching style (e.g. speed, volume, classroom presence)
- Workflow (e.g. pace, workload, technical issues, assignment/assessment feedback)
- Encouragement (e.g. things that work for you, welcomed adjustments)

This does not replace the student course evaluations taking place near the end of the course administered by the University. Students are encouraged to complete student course evaluations in addition to using the anonymous feedback tool.

Laptop Required for In-Person Learning

Students are expected to have access to a fully-charged laptop during in-person classes. Laptops may be used occasionally for quizzes and will allow students to code alongside the instructor for improved learning. **There are a limited number of electrical outlets in the classroom, so please remember to charge your laptop before class.** Alternative accommodations are available to students unable to meet this requirement for any reason. Please contact the instructor for further details.

Copyright Notice

Faculty members own copyright in their educational work at Texas A&M University, as stated in the Texas A&M University System Policy for Intellectual Property Management and Commercialization. Students are not allowed to post or share any materials created by a faculty member unless given permission by that faculty member. This includes but is not limited to homework assignments, homework solutions, exams, exam solutions, lecture notes and any other supplemental materials. **Any violation of this copyright policy could result in disciplinary actions as described in Student Rule 20.2: Procedures in Scholastic Dishonesty Cases and Student Rule 20.1.2.3.1.**

Texas A&M complies fully with the Digital Millennium Copyright Act ("DMCA"). Users of the Texas A&M network found to have engaged in repeated infringement of copyright are subject to termination of their network access and may be reported to the appropriate Dean or Human Resources officer for disciplinary action. Please see TAMU's Copyright Infringement Policies and Sanctions Notification for additional information.

Grading

- Course percentage performance (PP) yields grades as follows. A: $90\% \leq PP \leq 100\%$; B: $80\% \leq PP < 90\%$; C: $70\% \leq PP < 80\%$; D: $60\% \leq PP < 70\%$; F: $0\% \leq PP < 60\%$.
- The course grade will be based on attendance (5%), homework assignments (35%), quizzes (10%), a midterm exam (20%), and a final project (30%).
- **You must notify the instructor within one week after a score is posted to question a score.** If you do not notify the instructor of any issues with your score within that time, the instructor is under no obligation to re-evaluate the posted score and such requests will be handled on a case-by-case basis. Students should clearly explain why a regrade is appropriate. Requests lacking proper justification will not be considered.

Attendance

You are expected to attend the course in-person. The attendance component of your grade will be assigned as follows:

- One or fewer unexcused absences: 100%
- Two unexcused absences: 80%
- Three unexcused absences: 60%
- Four unexcused absences: 40%
- Five unexcused absences: 20%
- Six or more unexcused absences: 0%

To request consideration for an excused absence as defined in Student Rule 7, you must submit a request using this form prior to the day of absence. In cases where advanced notification is not possible, you must submit the request using the form by the end of the second business day after the last date of the absence. Documentation is required per Student Rule 7.3.

Here are some examples of reasons for absences that **will not** be excused:

- Preparing for an exam in another course.
- Completing assignments for this course or another course.
- Preparing for an interview taking place outside scheduled class time.
- Appointments you have the ability to reschedule.
- Transportation issues.

Homework Assignments

Homework assignments and due dates will be made available through Canvas. There will be regular assignments throughout the semester. **Homework assignments are to be completed independently unless explicitly indicated as a group assignment.**

Individual Assignments

Submissions must be your own work, not from outside sources, and consistent with the university rules on academic integrity. You are responsible for making sure that there is no reason to doubt that the work you hand in is your own. If only your name appears on an assignment, the assumption is that you have done the work yourself, fully and independently. I expect you to follow this policy scrupulously. Your chances for a good performance on the exams will be higher if you follow this policy.

Note that for homework you may use:

- Your textbook and notes from class.
- Your notes, homework, etc., from a related class that you took or are taking.
- References listed on the syllabus or mentioned in class.
- Discussion with the instructor or TA.
- Voluntary, mutual, and cooperative discussion with other students currently taking the class.

You may **not** use:

- Solutions manuals (printed or electronic) and copies of pages from solutions manuals.
- Solutions, notes, homework, etc., from students who took this class previously.
- Copying from students in this class, including expecting them to reveal their solutions in "discussion".

Be careful sharing your work with friends. **It is not appropriate to give your solutions to someone else to review and/or copy.** If I can determine that their solution is your solution, you will both get a zero on the assignment. This holds true even for ***partial*** sharing or if you did not know your solution would be copied.

Group Assignments

For group assignments, **all group members are expected to contribute roughly equal amounts of time and effort to the assignment.** The form of the contributions may vary according to each group member's individual roles, responsibilities, and expertise. However, there should not be extreme disparities in the amount of time and effort contributed to the assignment among group members. **Any issues raised to the instructor regarding lack of contribution from group members will be investigated on a case-by-case basis, and points may be deducted from individual grades accordingly.**

Good communication is key to ensuring each person understands their individual roles and responsibilities, dependencies across each member's workflows, and expected deliverables and deadlines. **You are expected to make every effort possible to attend group meetings, meet agreed upon deadlines, and coordinate with other group members depending upon your**

work to complete their work. Unresponsiveness and failing to meet deadlines previously agreed upon are unacceptable. If you encounter any difficulties within your group, please raise any and all concerns to the instructor in a timely manner, so they can be addressed before becoming problematic.

Quizzes

Quizzes may be administered during the scheduled class time or online with due date specified in Canvas. **In-class quizzes may be given without prior notice**, so please practice regular attendance, arrive on time to class, and enjoy the entire class period.

Midterm Exam

The midterm exam will be administered in class and may be proctored via Honorlock.

Final Project

The final project will be completed in groups and require both a group presentation and group report to be submitted. More details will be released after the midterm exam.

Proctoring

Quizzes and exams may be proctored using Honorlock, an online exam proctoring service. You do not need to create an account or schedule an appointment in advance. Honorlock is available 24/7, and all that is needed is a computer with a Chrome browser, a working webcam and microphone, and a stable Internet connection. After you verify your identity and scan your room, you can begin your exam. Honorlock will record you via webcam, as well as record your screen activity and web pages visited. Honorlock also has an integrity algorithm that can detect search-engine use, so please do not attempt to search for answers. Honorlock is FERPA compliant and uses securely encrypted protocols to save and view all test taker assets. Honorlock's certified proctors and your institution's faculty are able to review test session videos until deleted based on the school's retention policy. For answers to common questions on Honorlock proctoring, visit the TAMU Honorlock website and FAQ page or Honorlock's KnowledgeBase website.

Late Work

Late work is defined as submitting a deliverable (e.g. assignment, quiz, exam) after the established deadline or due date. Due dates will be strictly enforced. **For assignments submitted after the due date, students will receive a 30% late penalty deduction for assignments submitted within 7 days of the due date. Assignments will not be accepted beyond 7 days of the due date and will receive a grade of zero on the deliverable.** Please note that this does not apply if the late submission is due to an excused absence as defined in Student Rule 7.

Students should follow the procedure defined in Student Rule 7 when seeking an excused absence. Students seeking an excused absence must complete the form previously mentioned prior to the absence, if feasible, or by the end of the second business day after the last date of the absence. This notification must include an explanation of why notice could not be sent. The

student also is responsible for providing documentation substantiating the reason for the absence within three business days of the last date of absence.

If the reason is not approved, the absence is then unexcused. In this case, the above stated late work policy will be applied to any submissions received after the deliverable deadline. If the reason is approved, the absence is then excused, and the instructor will provide further instructions on making up missed deliverable(s). Work submitted by a student as makeup work for an excused absence is not considered late work and is exempted from the late work policy.

Incomplete

A temporary grade of I (Incomplete) at the end of a semester indicates that the student has **completed the course with the exception of a major quiz, final exam, or other work.** The instructor shall give this grade only when the deficiency is due to an excused absence or other cause beyond the control of the student.

Tentative Course Schedule

1. (2 weeks) Version control with Git and GitHub and deployment with Docker.
    - We introduce Data management and Git. We demonstrate Git and GitHub and expect the students to use these tools in their projects. When possible, the class will distribute, code, examples and data using Git.
    - Use existing Docker contains and build Docker containers to facilitate use and sharing of analytical projects.

2. (2.5 weeks) Relational databases: SQL
    - Cover principles of relational databases (design, schema).
    - We teach and use basic SQL including SELECT queries and related keywords, SQL joins, aggregation functions, and window functions.

3. (6 weeks) Introduction to data science Python features & libraries
    - We introduce and familiarize students with Jupyter notebooks.
    - Introduction/refresher to Python (basics, structures, functions).
    - The precise coverage will vary. Initially we look at pandas, numpy, matplotlib, seaborn, scikit-learn, and PyTorch.

4. (1.5 weeks) Non-relational databases: No-SQL
    - We discuss the principles of NoSQL databases.
    - We survey the different kinds of NoSQL databases.
    - We teach and demonstrate a key-value NoSQL database.
    - We use NoSQL within Python (key-value store).

5. (2 weeks) Parallel computing in Python with Dask
    - We introduce the basic Dask architecture that allows for parallelization (collections, task graphs, schedulers).
    - We cover low-level parallelization of existing code bases using the Dask Delayed decorator function.
    - We examine Dask collections for data storage and manipulation including Dask arrays and Dask DataFrames.
    - We create and use Dask collections and Dask Delayed for analyzing Big Data.

6. (1 week) Cloud programming and resources.
    - We survey cloud providers and the tools they offer concentrating on those of use to Data Scientist. We anticipate these to change over time.
    - If it makes sense to do so, use Python skills to interact with the cloud resources.

Frequency Asked Questions (FAQ)

1. Can I choose my group members for group assignments or choose to work alone?
   No. Students are randomly assigned to groups and are expected to make an effort to work effectively with all group members. Notify the instructor if there are issues working together effectively as soon as possible.

2. Is there a possibility for individual extra credit assignments?
   Out of fairness to all students, no individual extra credit assignments are given. All students will be given the same opportunity to earn points towards their final grades.

3. What is the format and level of difficulty for the exams?
   Exams are appropriately challenging for an upper-level undergraduate course. Exams are timed and will consist of numerous multiple-choice style questions (true/false, multiple choice, matching, fill in the blank, etc.).

4. What is the format and level of difficulty for the quizzes?
   Quizzes can vary in the format and level of difficulty. Some will require coding. Coding questions are challenging and require you to use multiple concepts introduced in the course simultaneously to solve practical problems. Do not expect these problems to be identical to those solved in class aside from a few numbers changed.

5. My programming background is not that strong. What should I do?
   Be acutely aware of your own strengths and weaknesses and work to resolve any deficiencies quickly. This may mean that you need to put in extra effort on certain topics. Please reach out if you would like any additional resources to help with particular topics.

6. What are some good strategies for working well within a group?
   Set deadlines for deliverables for each group member together as a group well in advance of the actual deadline to give yourselves buffer time. This way you have time to

review and revise the group submission before it is due. Also, this way your plan can absorb some delays here and there without impacting your ability to deliver on time.

Practice effective and regular communication within your group. For example, if you are unable to meet a deadline you previously set for your part of the submission, let your group know what's happening and when they can expect your work. Life happens. Talk about it with your group, so everyone is up to speed.

Put together a plan early on, so everyone can get to work and knows what to do and when parts are needed. Make every effort to meet agreed upon deadlines and deliver your contribution on time.

Raise up any issues within your group immediately to the instructor. The instructor can encourage and facilitate communication within a group and potentially avoid major issues. However, the instructor cannot easily address problems after the assignment has been submitted.

University Policies

Attendance Policy

The university views class attendance and participation as an individual student responsibility. Students are expected to attend class and to complete all assignments.

Please refer to Student Rule 7 in its entirety for information about excused absences, including definitions, and related documentation and timelines.

Makeup Work Policy

Students will be excused from attending class on the day of a graded activity or when attendance contributes to a student's grade, for the reasons stated in Student Rule 7, or other reason deemed appropriate by the instructor.

Please refer to Student Rule 7 in its entirety for information about makeup work, including definitions, and related documentation and timelines.

Absences related to Title IX of the Education Amendments of 1972 may necessitate a period of more than 30 days for make-up work, and the timeframe for make-up work should be agreed upon by the student and instructor" (Student Rule 7, Section 7.4.1).

"The instructor is under no obligation to provide an opportunity for the student to make up work missed because of an unexcused absence" (Student Rule 7, Section 7.4.2).

Students who request an excused absence are expected to uphold the Aggie Honor Code and Student Conduct Code. (See Student Rule 24.)

Academic Integrity Statement and Policy

"An Aggie does not lie, cheat or steal, or tolerate those who do."

"Texas A&M University students are responsible for authenticating all work submitted to an instructor. If asked, students must be able to produce proof that the item submitted is indeed the work of that student. Students must keep appropriate records at all times. The inability to authenticate one's work, should the instructor request it, may be sufficient grounds to initiate an academic misconduct case" (Section 20.1.2.3, Student Rule 20).

You can learn more about the Aggie Honor System Office Rules and Procedures, academic integrity, and your rights and responsibilities at aggiehonor.tamu.edu.

Americans with Disabilities Act (ADA) Policy

Texas A&M University is committed to providing equitable access to learning opportunities for all students. If you experience barriers to your education due to a disability or think you may have a disability, please contact the Disability Resources office on your campus. Disability Resources is located in the Student Services Building or at (979) 845-1637 or visit https://disability.tamu.edu. Disabilities may include, but are not limited to attentional, learning, mental health, sensory, physical, or chronic health conditions. All students are encouraged to discuss their disability related needs with Disability Resources and their instructors as soon as possible.

Title IX and Statement on Limits to Confidentiality

Texas A&M University is committed to fostering a learning environment that is safe and productive for all. University policies and federal and state laws prohibit gender-based discrimination and sexual harassment, including sexual assault, sexual exploitation, domestic violence, dating violence, and stalking.

With the exception of some medical and mental health providers, all university employees (including full and part-time faculty, staff, paid graduate assistants, student workers, etc.) are Mandatory Reporters and must report to the Title IX Office if the employee experiences, observes, or becomes aware of an incident that meets the following conditions (see University Rule 08.01.01.M1):

- The incident is reasonably believed to be discrimination or harassment.
- The incident is alleged to have been committed by or against a person who, at the time of the incident, was (1) a student enrolled at the University or (2) an employee of the University.

Mandatory Reporters must file a report regardless of how the information comes to their attention – including but not limited to face-to-face conversations, a written class assignment or paper, class discussion, email, text, or social media post. Although Mandatory Reporters must file a report, in most instances, a person who is subjected to the alleged conduct will be able to control how the report is handled, including whether or not to pursue a formal investigation. The University's goal is to make sure you are aware of the range of options available to you and to ensure access to the resources you need.

Students wishing to discuss concerns in a confidential setting are encouraged to make an appointment with Counseling and Psychological Services (CAPS).

Students can learn more about filing a report, accessing supportive resources, and navigating the Title IX investigation and resolution process on the University's Title IX webpage.

Statement on Mental Health and Wellness

Texas A&M University recognizes that mental health and wellness are critical factors that influence a student's academic success and overall wellbeing. Students are encouraged to engage in healthy self-care by utilizing available resources and services on your campus. Students who need someone to talk to can contact Counseling and Psychological Services (CAPS) or call the TAMU Helpline (979-845-2700) from 4:00 p.m. to 8:00 a.m. weekdays and 24 hours on weekends. 24-hour emergency help is also available through the National Suicide Prevention Hotline (800-273-8255) or at suicidepreventionlifeline.org.

Statement on the Family Educational Rights and Privacy Act (FERPA)

FERPA is a federal law designed to protect the privacy of educational records by limiting access to these records, to establish the right of students to inspect and review their educational records and to provide guidelines for the correction of inaccurate and misleading data through informal and formal hearings. Currently enrolled students wishing to withhold any or all directory information items may do so by going to howdy.tamu.edu and clicking on the "Directory Hold Information" link in the Student Records channel on the MyRecord tab. The complete FERPA Notice to Students and the student records policy is available on the Office of the Registrar webpage.

Items that can never be identified as public information are a student's social security number, citizenship, gender, grades, GPR or class schedule. All efforts will be made in this class to protect your privacy and to ensure confidential treatment of information associated with or generated by your participation in the class.

Directory items include name, UIN, local address, permanent address, email address, local telephone number, permanent telephone number, dates of attendance, program of study (college, major, campus), classification, previous institutions attended, degrees honors and awards received, participation in officially recognized activities and sports, medical residence location and medical residence specialization.

Other Policies

Tell Somebody

Oftentimes after a tragedy, people come forward with information and observations that, in retrospect, may have signaled a larger issue. This information when viewed collectively may be helpful in preventing tragic events and initiating assistance to an individual. Texas A&M University is committed to a proactive approach and needs your help.

As a member of this University community, if you observe any behavior that is concerning you may report the behavior using the online report form or by contacting one of the Special Situations Team members during business hours. The Special Situations Team is comprised of University faculty and staff charged with helping students, faculty and staff who are exhibiting concerning behavior.

This is not a system to be used for emergencies. If you are in an emergency situation that requires medical, psychological or police services, call 911.

Scholastic Dishonesty

It is the responsibility of both Students and Instructors to help maintain scholastic integrity at the university by refusing to participate in or tolerate scholastic dishonesty. Any violation of scholastic dishonesty could result in disciplinary actions as described in Student Rule 20.

Cheating: Intentionally using or attempting to use unauthorized materials, information, notes, study aids or other devices or materials in any academic exercise. Unauthorized materials may include anything or anyone that gives a student assistance and has not been specifically approved in advance by the instructor. Some examples of scholastic dishonesty are given here:

- *Improper Acquiring of Information*: Acquiring answers for any assigned work or examination from any unauthorized source. Working with another person or persons on any assignment or examinations when not specifically permitted by the instructor. Observing the work of other students during any examination both in class exams and take home exams.
- *Providing Information*: Providing answers for any work or examination when not specifically authorized to do so. Informing any person or persons of the contents of any examination prior to the time the examination is given.
- *Plagiarism*: Failing to credit sources used in a work product (homework, take home examination, paper, dissertation) in an attempt to take credit for the work of someone else. Attempting to receive credit for work performed by another person, including papers obtained in whole or in part from individuals or other sources.
- *Conspiracy*: Agreeing with one or more persons to commit any act of scholastic dishonesty.

- *Fabrication of Information*: The falsification of the results obtained from a research or laboratory experiment. The written or oral presentation of results of research or laboratory experiments without the research or laboratory experiment having been performed.
- *Violation of Computer Use*: Violation of any announced departmental or college rule relating to academic matters, including but not limited to abuse or misuse of computer access or information.
- *Aggie Honor System*: If you encounter students cheating or not abiding by university rules, then it is mandatory that you report the student to the Aggie Honor System Office: complete information is at aggiehonor.tamu.edu.