# This is the CCAC logo.

# DAT-202 – Data Analytics 2

# COURSE OUTLINE

**Class Section Time & Location**: Tuesdays 6-9:10pm on [Zoom](https://ccac.zoom.us/j/93363246341) & [Slack](https://ccac-data-analytics.slack.com)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Instructor: | Coral Sheldon-Hess | | Semester: | | Fall 2020 |
| Office Hours: | Mondays 2-5pm  Wednesdays 3-5pm | | Office Location: | | [Online](https://calendly.com/ccac-csheldon-hess) |
| Instructor Contact Methods: | **Email** (best way to reach me): [csheldon-hess@ccac.edu](mailto:csheldon-hess@ccac.edu)  *Note*: if you need to email me code, either paste it into the message text or save the file with a .txt extension  *Note*: emails will receive a reply within one business day  **Set up an appointment:** <https://calendly.com/ccac-csheldon-hess>  **Slack**: <https://ccac-data-analytics.slack.com/>, @coral (post your message in the **#dat-202-da2** channel, but feel free to tag me if I don’t reply immediately during class or office hours)  **Phone** (if you leave a message, I’ll hear it and reply within two business days): 412-369-4217 | | | | |
| Department Phone: | 412-369-4107 | Department Chair: | | [Rebecca Elinich DuPont](mailto:relinich@ccac.edu) | |

|  |  |
| --- | --- |
| Course Credits: | 3 |
| Pre- / Co-requisites: | DAT-201 |
| Course Description: | Building upon the principles set forth in prior coursework, students will engage in a comprehensive approach to the application of data analytics in the solving of business problems by employing the techniques frequently used in the discipline.  Emphasis will be placed on the different types of forecasting techniques such as sales, risk, retention and attrition as applied to a variety of industries. |
| Learning Outcomes  (from master course syllabus): | Upon successful completion of the course, the student will:   1. Differentiate the data needs for the different forecasting techniques. 2. Use appropriate database for specific forecast. 3. Implement data analytics forecasting. 4. Collaborate on using data for forecasting. 5. Use data visualization to illustrate forecasts. |

|  |  |
| --- | --- |
| Textbook & Materials | **Required textbook: *Regression Analysis with Python*,** Massaron & Boschetti, Packt publishing, ISBN: 9781785286315  **Required materials:** Students will need **access to a computer and internet** to attend synchronous sessions (which will also be recorded and made available on Blackboard) and to complete and submit assignments. To complete programming assignments, students will need access to a computer (not a tablet or mobile phone) on which they can install Slack and Anaconda 3 (both free).  Due to COVID-19, there is not a guarantee that CCAC’s campus computer labs will remain open all semester. This is also true for public libraries and other places where computer access outside of the home might be possible. Students should **make a plan for computer and internet access in case of renewed pandemic-related lockdown**.  Students will also need **a set of headphones** and **a microphone input to their computer**. (Headphones with integrated microphone should work on most machines. You can test your computer, headphones, and microphones at <https://ccac.zoom.us/test>) |

Course Policies & Procedures:

|  |  |
| --- | --- |
| Evaluation Plan: | As a lab-like course built around using data analytics tools to solve non-trivial, business-related problems, course assessments in DAT-202 are based on fully-baked student work products.  The instructor provides incremental feedback to students during the course of the module's individual project work time--often called formative assessment. Small misunderstandings or trouble spots that emerge inside a module can be ironed out before they impede the larger learning goals of the component. After all modules are mastered and a final project completed, the instructor offers additional, formal feedback concerning the project's alignment to its design specifications is provided.  Students complete the following steps in in advance of their presentation and feedback session for their culminating project:   * Project design specifications * Project flow diagram adjusted to reflect actual implementation * Thoughtful responses to "heart-of-the-matter" questions   ***Using design criteria alignment in place of rubrics***  The best assessment tools are those with which the students directly engage in creating and using. This can take the form of a class-generated project rubric, for example. As students create assessment criteria prior to implementing a project, the resulting work is both more likely to align to the assessment criteria and meaningfully assist students in completing their work. When that rubric is then used by the students to assess their own work, valuable mental processes are underway which tend to naturally improve skill and confidence.  Rubrics are widespread and useful tools for many types of student work outside of the technical design realm. In a technical class, such as this data course, the process of assessing student code against initial design requirements often organically takes the place of rubric-based assessment without displacing its generic value as a teaching tool.  ***Mapping project performance to course letter grades***  The following table serves as a possible correlation guide between module and component project assessment and the formal course letter grades instructors assign to each student at the conclusion of the semester:  A - Independent practice for **each model is completed and documented.** Culminating projects for each component meet all specified design criteria. Component reflections show **evidence of synthesis**with other technical learning domains.  B - Independent practice for each module has been **attempted but not consistently documented** to reveal command of the code. Culminating projects for each component meets some but not all design criteria. Component reflections show **moderate thought, limited to current learning topics.**  C - Independent practice for **1/2 to 2/3 of modules has been attempted** but not consistently documented. Culminating projects for each component **meets some but not all** design criteria. Component reflections show **low levels of thought** relative to A and B work.  D - Independent practice for **less than 1/2 of modules** has been attempted but not consistently documented. Culminating projects for each component meets few, if any design criteria. Component **reflections are incomplete.**  F - Independent practice for 1/4th or fewer of modules has been attempted and not consistently documented. **Culminating projects were not meaningfully attempted.** Component reflections were not attempted.  ***Documenting work done outside of class***  Each student is expected to document the time they spend on their studies outside of classroom time. This documentation should serve as a self-assessment tool, but may potentially contribute to peer- and instructor-assessment, as well. The format of the documentation is left up to the individual student, though a  spreadsheet with the following categories would be a good starting point:   1. Background reading, including documentation and tutorials 2. Project work (hands on keyboard) 3. Design & pondering, sketching out solutions 4. Collaboration   Especially when projects are completed as a group, each individual will be expected to contribute and to be able to document their own contributions. |
| Attendance & Tardiness: | If you need to miss our synchronous session occasionally, that is all right; they will be recorded and posted on Blackboard within a day. Attending them live is a *really good idea*, because questions *will* come up as we cover material; having the chance to ask and get clarification in real-time is invaluable. |
| Test/Quiz Makeup: | N/A – no tests or quizzes |
| Technology Use: | Much of the practice of data analytics involves wrangling the various software products we need to do our jobs. As such, students are welcome to use their own machines to complete their work. Some class time will be spent on tool-wrangling, but students will also be expected to spend time outside of class on software installation and configuration. |
| Academic Honesty: | Copying or failure to turn in your own work will result in a score of a zero on the assignment, notification of the department chair, and a required meeting with the instructor or department head. Additional consequences could include failing the course and any other academic consequences the instructor and the department chair deem appropriate.  **If you use external sources for help on your assignments, you are required to document those sources.**  This is a higher-level data analytics course. Consulting documentation and outside sources is expected and welcome, but citation of those sources is also vital. |
| Available help: | Tutoring is available via Zoom for all CCAC students – more information at [https://ccac.edu/Learning\_Commons.aspx](https://ccac.edu/Learning_Commons.aspx%20) & [https://ccac.edu/Tutoring/](https://ccac.edu/Tutoring/%20)  Up to 10 hours of online tutoring are available through Smarthinking (accessed via <https://ccac.edu/Tutoring/>)  If you’re having trouble with the course, you’re definitely welcome to use tutoring, but please also feel free to use my office hours. You can make appointments with me via Calendly: <https://calendly.com/ccac-csheldon-hess> |
| Other Policies and Procedures: | In this class we will function as a team. We will work individually and together to build necessary skills that are required to work in a data analytics environment. Some of the learning tools we may use include lectures, readings, coding projects, discussions, and workshops.  Just like on a real data analytics team, our goal is for all of us to succeed. We are a community of learners, responsible for one another’s success.  Students should prepare before each class doing the readings that will be discussed during the next class period and completing any lab activity and/or homework that is required from the previous class period.  That all said, we are all humans, living through unprecedented times. Your professor is a human and recognizes that so are you. If you aren’t going to be able to turn something in on time, let your professor know *before the day it’s due*, and they will do whatever they can to work with you.  If you or a family member get sick, if your job’s hours change, or something else happens that’s going to put you in danger of not finishing the course by December 12, please reach out—some deadlines can be modified, and, very worst case, we can set up a plan for an “Incomplete” grade, which will allow you to finish the course later. |

All students are expected to read and comply with the policies and regulations set forth in the CCAC Student Handbook, including without limitation the College’s policies regarding academic and behavioral conduct, the procedures for requesting an accommodation based upon a disability, pregnancy or pregnancy related condition, or a religious observance, and for reporting unlawful discrimination and harassment.

The Student Handbook is available to view and download from the College’s website at the following URL:  <https://www.ccac.edu/academic-rules-and-regulations/rules-and-regulations.php>.

The full text of the College’s *Policy Manual,* *Administrative Regulations Manual*, and the Civil Rights Complaint Procedure can also be viewed and downloaded at:  <https://www.ccac.edu/president/policies-and-regulations.php>; <https://www.ccac.edu/diversity/title-IX.php>; and <https://www.ccac.edu/diversity/notices.php>.

Information concerning the process and documentation required to request a disability-related accommodation can be obtained by contacting the campus’ Office of Supportive Services for Students with Disabilities (OSSSD) or by visiting the OSSSD information page at <https://www.ccac.edu/supportive-services/suppotive.php>.

Students are reminded that they can access their course information and CCAC email account, the CCAC Academic Calendar (including add/drop/withdrawal deadlines), the Student Handbook, the College’s Incident Report form, and many other College services through the MyCCAC portal at:  [https://my.ccac.edu](https://my.ccac.edu/).

| Course Plan: | | |
| --- | --- | --- |
| Class Week/Date | Topics / Learning Activities | Assignments / Homework | |
| Week 1  *9/1* | Welcome and introductions  Zoom and Slack info for the semester  What even is forecasting | Read chapter 2 of Brownlee  Readings on de-anonymization / PII | |
| Week 2  *9/8* | De-anonymization & privacy discussion  Planning out the semester | Readings on bias in big data | |
| Week 3  *9/15* | Simple linear regression – the concept and also implementation in Excel | Video: [Practical privacy-preserving machine learning in Python](https://www.youtube.com/watch?v=NUk6QN02UxQ) (PyCon) | |
| Week 4  *9/22* | Discussion of privacy-preserving ML  Video:   * K-nearest neighbors - Michael * Bayes - Drew | Video: [Implementing Ethics: Developing Trustworthy AI](https://www.youtube.com/watch?v=vPokIvli8yk) (PyCon) | |
| Week 5  *9/29* | Discussion of trustworthy AI  Going over knn & Bayes assignments  Video:   * Data cleaning – Kayla * Open Refine - Zach | HIPAA readings tbd | |
| Week 6  *10/6* | HIPAA discussion  Going over data cleaning & OR assignments  Video:   * ARMA – Brandyn * ARIMA – Jason * scipy – Jinrong | FERPA readings tbd | |
| Week 7  *10/13*  *Midterm grades* | FERPA discussion  Going over assignments on ARMA, ARIMA, & scipy  Video:   * Neural networks – Vajiheh * Exponential smoothing – Eric * Using forecasting to address disparities in the US - Jill | GPDR, readings tbd | |
| Week 8  *10/20* | GPDR discussion  Going over assignments on NN, ES, & forecasting for social good  Broad (very broad) ML overview, with a focus on regression analysis | Data analysis for social good, readings tbd | |
| Week 9  *10/27* | Data analysis for social good  Dimension reduction, maybe?  Stochastic models, Markov models? | Ethical implications of using invalid data, readings tbd | |
| Week 10  *11/3* | Ethics of data validity  Survey instruments, avoiding bias in data acquisition | Informed consent & IRBs, readings tbd | |
| Week 11  *11/10* | Informed consent & IRBs  Algorithm design for ML? (Maybe!) | [Prophet reading 1](https://research.fb.com/prophet-forecasting-at-scale/)  [Prophet reading 2](https://facebook.github.io/prophet/docs/quick_start.html#python-api)  [Time series forecasting with Prophet](https://machinelearningmastery.com/time-series-forecasting-with-prophet-in-python/)  Ethical implications of p-hacking, multiple testing, readings tbd | |
| Week 12  *11/17* | p-hacking, multiple testing  Forecasting with Prophet |  | |
| Break  11/23-11/29 | Thanksgiving Break | No new assignments over break, but you may want to work on your project! | |
| Week 13  12/1 | Final project work time | Finish your final project and turn it in by 6pm on 12/10/20 | |
| Week 14  *12/8* | Final Projects | Final project sharing over Zoom | |

Course Outline Corrections:

During the semester/session, reasonable changes to the course outline may be academically appropriate. Students will be notified of these adjustments by the instructor in a timely manner.