CARLOW UNIVERSITY

Course Information	**DTAN-310-DA: Big Data and Data Systems** Spring 2023 | Main Oakland Campus Monday/Wednesday 3:30 PM - 4:45 PM | 1/9/2023 - 5/5/2023 A J Palumbo Science & Tech, 202 Onsite-Lecture
Instructor	Eric Darsow
Contact Information	ecdarsow@carlow.edu 412.894.3020 (landline; preferred on non-class days) 412.636.42356 (mobile; no SMS; only useful on class days on campus)
Department Chair	Dr. Ericka Mochan Email: edmochan@carlow.edu | Phone: (412) 578-2053 | Office: AJP 307
Office Hours:	Mondays & Wednesdays: 1:00 pm to 2:00 pm In AJP classroom 202 if available or at table area on the second floor
Textbook	None required; Readings assembled for individual modules
Course Description	This course provides an overview of big data and the types of analytics used to process this data, as well as the associated technical, conceptual, and ethical challenges of dealing with big data. Advantages and disadvantages of big data research are discussed using real-world examples and case studies. The course includes hands-on exercises working with big data in Python. Prerequisite: DTAN 150. *[From Dr. Mochan's course description]*
Course Goals & Objectives	Data analytics can be described as the process of extracting information and drawing conclusions from complex data. Big data analysis is a subset of analytics that features very large data sets (data with thousands or millions of rows/columns, which requires a large amount of storage space on the computer to save and work with). Students will exhibit big data skills by: • Downloading, reading, and saving large datasets in an efficient way • Cleaning and organizing raw data • Visualizing data with appropriate graphs and tables • Using data to predict future events and make decisions • Drawing statistically sound conclusions from the data and communicating these results to a lay audience • Discussing common ethical issues with case studies in big data **Upon completion of this course, a student will be able to** • Understand the characteristics of big data, including its advantages and disadvantages. • Learn basic techniques for loading, reading, processing, and analyzing big data on an average computer. • Apply Python packages Pandas and sci-kit-learn to analyze big data sets. • Understand the limitations of Python in processing big data. • Evaluate existing big data processing platforms and tools, and understand the advantages and disadvantages of each. • Describe the MapReduce process and how it applies to big data. • Analyze and describe real applications of big data and the ethical issues associated with the data, including privacy, ownership, transparency, and consent . *[From Dr. Mochan's course objectives]*

Course structure	*During this exclusively project-oriented course* student will cultivate fluency in designing data visualizations appropriate to the structure of that data and the relevant inquiry goals. We will practice our module topics by designing and implementing "mini-projects" undertaken over a week or two and shared for comment and critiqued by peers during class time. Students will assimilate their skills by choosing one of their module-level projects to build into a fully-formed ("fully-baked") culminating inquiry for sharing with the class during the final exam period. **Weekly learning cycle explained:** Mondays of each week *will be for introducing new materials and skills*. Students should plan on dedicating a few hours out of class between Monday and Wednesday's class sessions to practice the module basics through the practice exercises we'll introduce in class. Wednesdays: We'll review the learning process so far on that week's topic by sharing bits of our practice from out of class. Then we'll deepen our engagement with the topic by applying the tools to novel data sets, such as those related to your domain of interest. During the last half of the week and weekend, students apply the week's content to a novel "mini-project" to share on the next Monday as a bridge into our next topic. Some multi-week modules will elongate this pattern over two or three weeks with the projects requiring a more substantial engagement. **Attendance tracker:** Each student is issued a pocket folder that contains their own attendance sign-in sheet, which is a table with a row for each class session and if present, students sign on that line and report any work completed. Strive for 80% attendance and makes notes of reasons for absences. **Completion targets & deadlines:** The major deadline for this class is the sharing of a fully-baked final project during the final exam session (in lieu of a written exam). Intermediate project deadlines during the course should be attended to with conscientious focus and work products brought to class for sharing. Strive to create semi-final work products for each mini-project but do not fear immersing yourself in a learning stream that means outputs remain preliminary during formative learning. **Sharing work with peers:** Sharing one's work at all stages of the design and implementation process is essential to this project model's success. We will do so with an emphasis on giving credit and access instructions when we getting help from books, forums, courses, peers, etc. Comment your code carefully with credits as they are relevant. **We value in-process work:** Given the iterative nature of technical learning, students are invited to be comfortable--even confident--bringing their data project work to class for sharing even if roadblocks, bugs, or intense questions are encountered. Your instructor will strive to cultivate a learning environment in which time in class is collaborative at all levels of the learning process. Students are encouraged to boldly communicate the rawness and messiness of their own trajectories for the benefit of the entire class. You are encouraged to help each other through the dark and sometimes dreary parts technical undertakings with large, messy data sets.
Letter grades	Given the diversity of backgrounds and skills of students in this course, letter grades will be developed collaboratively between students who share documentation of work completed and the instructor who sets grading norms. This will occur both at mid-term and again for formal registration with Carlow at the beginning of our final exam session. Both grade proposals consist of a proposed letter grade followed by list or short description of verifiable statements of work products completed, skills acquired, and course contributions made. Grade proposals

	are for reporting the actual state of work at that exact time, not in an aspirational way. Student letter grade proposals are written privately by each student on a 3x5 card. The process is public, but anonymous: you'll be assigned a unique ID number to print on your card which is only mapped to your name in the grade book. This process allows for calibration of one's own recommendations given the overall experience of students in the course. To explore artifacts of this process in action, review past grade proposals posted on your instructor's archive server from CCAC Data Analytics and Java programming courses: https://technologyrediscovery.net/coursesGen/letterProposals_sp19.html **Benefits of student-driven grading** Your instructor has found that this student-driven grading approach allows students of a diversity of incoming skill levels to focus in a relatively low-stress fashion on acquiring skills at the *level optimized for their particular tier of data analysis competencies.* For example, students already fluent in computer programming languages ripe for creating data visualizations can pursue additions to their already developed coding skills while more novice students can apply course concepts by adapting and tinkering with pre-built code modules with higher levels of peer, tutorial, and instructor support. Students submitting grade proposals should use the following letter-grade based guidelines in calibrating their own proposals, using + and - designations when category lines are blurred. **Letter A Grades:** Reserved for student who can demonstrate consistent and significant effort to deeply engage with the vast majority of our topic modules through skills practice that yields substantive work products. Further, students proposing A grades should be able to demonstrate some form of meaningful contribution to our class community, such as by consistently peer-tutoring, thoughtful peer work reviews, active class engagement, etc. Students proposing A grades should have a "fully-baked" final project complete at the time of proposal. Class attendance has been very good (~90%). **Letter B Grades:** Appropriate for students whose efforts in the course have been generally consistent over the term (e.g. not crammed into the last half or third of the course). Students have made a good-faith efforts made to engage with most of the course's module contents with attendance in the 80% range. Students have made a solid attempt at fully-baking a final data inquiry project for sharing with the class during finals. **Letter C Grades:** Appropriate for students whose course engagement has been inconsistent throughout the term and whose work products reveal only partial engagement with some of the course modules. C letter grades often reflect an incomplete culminating project. **Letter D Grades:** For students whose engagement in the course has been extremely minimal, with attendance 60% or lower. Students did not meaningfully attempt a culminating project **Letter F Grades:** Reserved for students whose engagement with the course has been near zero. **Disputes:** During your instructors 5+ years of engaging with letter grading using this approach for both introductory and upper-level project courses in both data analytics and computer programming, divergences between student and instructor letter grades have *always been resolved* through conversation and review of actual work products along with the chance to submit remedial work to close gaps in aspirational versus tangible work. In cases where instructor and students cannot agree on an appropriate letter grade, the Carlow Data Analytics/Math Department Chair will be asked to review the work products of the student asking for review in context of work products created by other students and their grade proposals.

Course schedule

Week	Topics & Milestones
Wk 1: Mon-9-JAN Wed-11-JAN	Big data analytics as an immense scientific and commercial force for discovery, inquiry, and planning. Classification tree of "big data" technologies • Massive hand-held CPU cycle potential has allowed artificially intelligent code to run even on personal and handheld computing devices ◦ Pattern recognition locally & with server support • Multi-core processing vs. explicitly parallel processing
Wk 2: (MLK Jr.) Wed-18-JAN	Families of big data • Environmental & biomechanical sensor data ◦ (human patient status, anaerobic digestion reactions, weather, aquifer levels, sewage treatment efficacy, water quality, air quality) • Machine operation log data: ◦ Click patterns on user content ◦ Location pings, device type registration, application/OS/adapter registration and logging ◦ User login patterns • Social science data: ◦ Study data warehouses ◦ UofPitt All of Us biological warehouse Tool setup: • Python and R environment configurations
Wk 3: Mon-23-JAN Wed-25-JAN	Relational data crash course: Table design from flat file structure Basic INSERT and basic SELECT More advance JOIN styles
Wk 4: Mon-30-JAN Wed-1-FEB	Python core data type and iterating Python libraries (numpy, pandas, matplotlib) Connecting Python to a database: SQLITE
Wk 5: Mon-13-FEB Wed-15-FEB	Explanatory data analysis • Pandas: .describe() • Basic plots, paneling plots of correlations
Wk 6: Mon-13-FEB Wed-11-JAN	Longitudinal data and forecasting models (starting with running averages)
Wk 7: Mon-20-FEB Wed-22-FEB	Working with US Census data and visualizing error rates at various levels of data grouping
Wk 8: Mon-27-FEB Wed-1-MAR	Accessing data over a network: API access vs. scraping Essential API authentication
Wk x: Mon-6-MAR Wed-8-MAR	Managing data on distributed file systems Cloud computing versus Virtual Private Servers
Wk 9: Mon-13-MAR Wed-15-MAR	The ethical aspects of big data systems: • European general data protection vs. US wild west • Third part data aggregators • Government oversight capability of AI-based data systems

		○ What if the companies themselves don't know what data they have or do not have because they let machines build their own predictive models.
	Wk 10: Mon-20-MAR Wed-22-MAR	Big data case study: network forensic data
	Wk 11: Mon-27-MAR Wed-29-MAR	Big data case study: climate change at varying scales
	Wk 12: Mon-3-APR Wed-5-APR	Big data case study: Facial recognition on your own laptop
	Wk 13: Mon-10-APR Wed-12-APR	The concept of MapReduce and Hadoop Solving very large problems in very small parts
	Wk 14: Mon-17-APR Wed-19-APR	Project design and work time
	Wk 15: Mon-24-APR Wed-26-APR	Final project preview presentation and peer feedback
	Finals week 1-5 MAY	Fully-bake final projects and prepare to share and share

Unified Carlow Policies

STUDENT SUPPORT SERVICES AND POLICIES:

The Center for Academic Achievement (CAA), 4th floor University Commons, offers free in - person
tutoring for improving writing skills and understanding course content. We also offer academic coaching for time management and learning skills. Make an appointment through https://carlow.mywconline.com at least 48 hours in advance, or call 412-578-6146.

Cancellations can be made online within 8 hours of the appointment time. For last minute cancellations, please call 412-578-6146. Students can also access “CAA Resources” under Groups on CelticOnline for guides on college survival skills, reading, learning strategies, math, and writing/research. Remember: the purpose of tutoring is to enhance independent learning, so tutors do not “edit” your papers or do your homework for you. Students are active participants in the tutoring experience.

STUDENTS WITH DISABILITIES POLICY

Carlow University makes every effort to provide reasonable accommodations for students with disabilities. This includes individuals with physical disabilities, learning disabilities and mental health disorders who meet the definition of disability under the Americans with Disabilities Act. Students with disabilities have the same responsibility as other stud ents to meet the University's academic, technical, and behavioral standards and to follow the University's general policies and guidelines regarding standards of conduct. Students who plan to request accommodations should contact the Disabilities Services Office at the beginning of each semester since accommodations cannot be granted retroactively. To determine whether you qualify for accommodations, or if you have questions about services and procedures for students with disabilities contact:

Jacqueline M. Smith
Disabilities Services Office, University Commons, 4 th floor
Phone - 412.578.6257 (Office line); 412.578.6050 (Direct line)
Fax - 412 578.2027

dso@carlow.edu

CARLOW UNIVERSITY TEMPORARY DISABILITY POLICY

Carlow University values each student and is invested in encouraging his or her academic success in line with the Mercy mission "to respond reverently to God and others; and to embrace an ethic of service for a just and merciful world." In keeping with the mission, the University has chosen to offer assistance to those with temporary conditions such as short term illnesses, injuries, or other temporary medical conditions. While the University is not required to provide such support under the Americans with Disabilities Act, some assistance may be arranged via the Disability Services Office (DSO). Each situation will be reviewed; however, the office cannot guarantee that services will be provided.

In order to determine if a student with a temporary condition may receive some assistance via the DSO, he/she should contact the office at 412 578-6257. The student will be asked to meet with Jackie Smith, Disabilities Services Representative, and to provide the requisite documentation of his/her condition. Mrs. Smith will review the documentation and may consult with the student accommodation committee to determine what, if any, assistance may be provided. All documentation will remain confidential.

ACADEMIC INTEGRITY POLICY

Carlow University aims to educate and challenge students to reach their highest potential by guiding students along a path of honesty and integrity throughout their intellectual pursuits. Students are thus expected to uphold the highest standards of academic integrity. Forms of academic misconduct include (but are not limited to):

Cheating—providing or receiving inappropriate assistance on any coursework.
Plagiarism—submitting another's work as one's own; not properly citing sources, using exact wording
without quotations or proper attribution, paraphrasing without proper citation, or improper paraphrasing; attributing citations to inaccurate or misleading sources.
Self-plagiarism—unauthorized use of one's own work or part of a work, either from the same course or from another course, in more than one assignment.
Academic deceit—use of false or altered information or withholding information critical to the processes of the University; providing false information or documentation with the intent to obtain an exemption, extension or exception to one's coursework; signing other students into classes or on group reports.
Fabrication of data—using falsified or fabricated data, forgery, or unsanctioned documents for research or other coursework.
Interference with other students' learning or achievement—sabotaging (including failing to contribute to) group projects or laboratory work, disrupting in -class work, altering computer files or online posts, or making educational materials unavailable to others.
Unauthorized acquisition or exchange of coursework—purchasing, borrowing, stealing, or otherwise obtaining material with the intent to use as one's own coursework; selling, lending, or otherwise offering one's own coursework to others with the intent of allowing the recipient to use the work as one's own; obtaining a copy of one's own completed tests and exams (either a physical copy, an electronic image, or a screenshot) without explicit permission from the course instructor.
All violations of Carlow's academic integrity policy will be kept on permanent record. Serious or multiple violations will be forwarded to the Academic Integrity Committee for a judicial hearing. It is the student's responsibility to become familiarized with Carlow's Academic Integrity Policy. The full policy can be found in the Course Catalog.