DSCI 521: Data Analysis and Interpretation

Course Syllabus

Credits: 3-hour lecture [3 credits]
Winter 2021
Short Title: Analysis and Interpretation

General Information

Course Instructor: Jake Ryland Williams
Instructor Contact Information: jw3477@drexel.edu
Office Hours, Location: Wednesdays, 5pm

Student Learning Information

Course Description

Introduces methods for data analysis and their quantitative foundations in application to pre-processed data. Covers reproducibility and interpretation for project life cycle activities, including data exploration, hypothesis generation and testing, pattern recognition, and task automation. Provides experience with analysis methods for data science from a variety of quantitative disciplines. Concludes with an open-ended term project focused on the application of data exploration and analysis methods with interpretation via statistical, algorithmic, and mathematical reasoning.

College/Department: College of Computing & Informatics
Repeat Status: Not repeatable for credit
Restrictions: None
Prerequisites: DSCI 501 (pass the test to waive)

Course Purpose Within a Program of Study

This course provides an application-oriented introduction to quantitative reasoning for data science. Hands-on experience is provided with a number of analytic techniques and evaluation methods in the context of their quantitative foundations. Via implementation, an understanding for method function, applicability, setup, and interpretation is provided under mathematical frameworks.

This course is a core course in the Data Science Masters program.

Statement of Expected Learning

The course objectives are to:

- observe and explore a variety of quantitative methods for data analysis;
- understand methods' evaluation techniques to interpret their output;
- implement and evaluate methods to gain technical experience with data; and
- reproducibly execute an analytic project and represent/communicate its results faithfully.

As learning outcomes, students completing this course should be able to take data and a task and identify methods for analysis, reproducibly implement them by using widely-available tools, reason quantitatively to interpret their results, and conduct evaluations for the faithful communication of output.

Course Materials

Lecture Notes

Students will be responsible for reading these primary course materials in depth, on their own. These lecture notes consist of a collection of interactive Jupyter notebooks, which may be found on the course blackboard website.

The course lecture notes are broken down into the following topics, which roughly correspond to a week of content each. More information on course scheduling can be found in that section, below.

> Chapter 0: System configuration, programming, and numeracy
> Chapter 1: Processing numeric data
> Chapter 2: Feature engineering and language processing
> Chapter 3: Exploratory data analysis and visualization
> Chapter 4: Networks, metrics, and applications
> Chapter 5: Data transformation and analysis of functions
> Chapter 6: Counting and probability for quantitative frameworks
> Chapter 7: Approaching regression and classification problems
> Chapter 8: Model pipelines and comparisons
> Chapter 9: Neural network models

Video Recordings

As this course will proceed entirely online, we will be using recordings from Dr. Williams' past, face-to-face class. These recordings will be distributed through Blackboard. The recordings have a length of about 170 minutes.

Weekly exercise

There will be weekly coding exercises released on Bb Learn.

Office Hours

Office hours will be in a question & answer format in which students are required to bring any of their questions related to the course content, assignments and course project. Office hours will be regularly scheduled (TBD), via Zoom and also with availability by appointment (as needed).

Texts, Readings, and Resources

Note: all text readings are supplemental to the course lecture notes and will be assigned on a weekly basis. Python Data Science Handbook is available free of charge on Github:

> • https://jakevdp.github.io/PythonDataScienceHandbook/

and The Data Science Handbook and Data Science from Scratch are available to Drexel students through the University Libraries. Specific text information is as follows:

- Data Science from Scratch (DSFS): First Principles with Python. ISBN: 978-1492041139, O'Reilly Media, 2019
- Python Data Science Handbook (PDSH). ISBN: 978-1491912058, O'Reilly Media, 2016
- The Data Science Handbook (TDSH). ISBN:9781119092940, John Wiley & Sons, 2017

Required and Supplemental Materials and Technologies

Note: instructions and discussion of the following materials and technologies are provided in Chapter 0 of the course lecture notes (below). Students are expected to have the following by the start of the first week:

- A command line environment with Python (version 3) installed
- The Jupyter notebooks interactive development environment

Assignments, Assessments, and Evaluations

Graded Assignments and Learning Activities

Quizzes: Weekly preparation assessments must be completed before class and will each count for a small percentage of your overall grade. Quizzes will be available with unlimited attempts, are primarily intended for students as a preparation self-assessment for synchronous sessions, and will become available one week prior to them on Blackboard (your week 1 quiz should already be available).

Homework: Structured assignments will be distributed according to four topic areas:

1. Text analysis and feature engineering
2. Network and exploratory data analysis, including visualization
3. Probabilistic modeling and prediction
4. Machine learning and regression

These assignments will be composed in a modular fashion (11 modules, total), with each module/problem worth about 50 points (524 points, total). Modules may be submitted optionally in groups of up to four, consisting strictly of classmates from the same section. Critically important assignment submission information is as follows:

- Submissions will only be accepted through Blackboard (no email submissions)
- For each module, submit only your module-X.ipynb file (Jupyter notebook)
- All relevant submissions will be marked after their Answer Keys have been released
- Any grading disputes must be resolved within 5 days of receipt (contact grader, cc instructor)
- All team member names (even if submission is individual) must be present in the designated header area in all module notebooks, subject to a 2-point non-compliance penalty.
- All team member names (even if submission is individual) must be present in the Blackboard submission comments, subject to a 2-point non-compliance penalty.
- All group submissions must be submitted as identical, by each group member, subject to a 2-point non-compliance penalty.
- If you receive help for any question in any individual module, indicate the person's name in both the submission comments and within, subject to a 2-point non-compliance penalty.
- If you print an unreasonable volume of output in your notebook the instructor may remove up to 2 point per module section, regardless of code correctness.

Project: One open-ended group assignment will have two phases:

1. Data Set Identification, Motivational Exploration, and Proposed Methods Implementation
2. Methods Implementation, Evaluation, and Interpretation, with Documentation and Dissemination

The Project Proposal will be graded based on the following rubric:
   a. Forming a team (include in the project proposal team's members self-identified skills, and individual contributions).

b. Identification of an existing data set and analytic subject of interest
c. Discussion of analysis stakeholders, use, or intention.
d. Discussion of the plan for performing the analytic experiment.
e. Presenting a sample of experimental data, and exhibiting potential tools.
f. Submission of the project proposal.

The Project Implementation will be graded based on the following rubric:
g. Project results
h. Documentation (including Jupyter notebooks, ReadMe file, Python code, data dictionary and other supporting documents)
i. Submission (submit to OneDrive request file link, one copy per group)
j. Presentation (has one slide/paragraph of workload distribution)
k. Evaluation & comment on other groups project presentation

Teams for Assignment and Project

The assignments and project are team effort.
- Each team can have up to 4 members for both assignment and project. Student is allowed to work alone in a one-person group.
- Members can always change for each assignment.
- Once project proposal is submitted, it is recommended to not change members in project group. If team members change, a written request must be submitted to the instructor.
- Each team submits a single completed project proposal and a single project zip file for the final project.
- Each team will be asked to present the project at the end of the term.

Team Member Project Evaluation

- All students on a single team initially receive the same grade. However, each team member will evaluate the performance of every other member in his/her team. The instructor reserves the right to adjust the grade based on the team evaluation results.
- All students of a team must have their names listed clearly on the front page of the reports. Each team's work must be unique – one team cannot collaborate with another team.

Grading Matrix

Students will not receive letter grades for individual assignments. Grades are calculated as:

Project:	30% (10% Proposal, 20% Implementation)
Quizzes:	10%
Homework:	60%

Total:	100%

Grade Scale

The following scale will be used to convert points to letter grades:

Points	Grade	Points	Grade	Points	Grade
97-100	A+	84-86.99	B	70-73.99	C-
94-96.99	A	80-83.99	B-	67-69.99	D+
90-93.99	A-	77-79.99	C+	64-66.99	D
87-89.99	B+	74-76.99	C	0-63.99	F

Note that the instructor may revise this conversion if/when necessary.

Course Schedule

The course's schedule follows the lecture notes at roughly one week per chapter with the expectation that students will configure systems and review or work through the numeracy and processing fundamentals in **Chapter 0**. Some chapters may extend over multiple weeks, depending on the section's pace. **Week 10** and the regularly scheduled **final exam** period (the same day of week and time of the regularly scheduled class meetings, but during exam week) are reserved for final project presentations. Please observe the following (tentative) schedule; weekly activities may change depending on the pace of class.

- Week 0:
  - DSFS: Chapter 1
  - TDSH: Chapters 1–2
    - Homework: configure system
- Week 1:
  - DSFS: Chapters 2 & 4
    - Project: team formation; begin Phase 1
    - Homework: begin Assignment Group 1
- Week 2:
  - DSFS: Chapters 10 (pgs. 121–133) & 20 (pgs. 239–244)
  - TDSH: Chapter 16
    - Project: continue Phase 1
    - Homework: continue Assignment Group 1
- Week 3:
  - DSFS: Chapters 3 & 5
  - TDSH: Chapters 5.1–5.12
  - PDSH: Chapters 4.01–4.02, 4.04–4.06, 4.08–4.09
    - Project: continue Phase 1
    - Homework: begin Assignment Group 2; complete Assignment Group 1
- Week 4:
  - DSFS: Chapter 21 (pgs. 255–256, 264–266)
    - Project: begin Phase 2; complete Phase 1
    - Homework: continue Assignment Group 2
- Week 5:
  - DSFS: Chapter 8
    - Project: continue Phase 2

- o Homework: continue Assignment Group 2
- Week 6:
  - DSFS: Chapter 6
  - TDSH: Chapters 6, 18.0–18.3, 18.5–18.7, 19.0–19.2
    - o Project: continue Phase 2
    - o Homework: complete Assignment Group 2
- Week 7:
  - DSFS: Chapters 14, 15 (pgs. 171–183), & 16 (pgs. 189–192, 196–200)
  - TDSH: Chapters 8.1–8.4, 8.5.5, 10.6, 10.7 (through 10.7.2), & 11.0–11.6
  - PDSH: 5.06 & 5.09
    - o Project: continue Phase 2
    - o Homework: begin Assignment Group 3
- Week 8:
  - DSFS: Chapters 11, 13, 17 (pgs. 201–203 & 211) & 19
  - TDSH: Chapters 8.5.1–8.5.4, 8.5.7
  - PDSH: 5.01–5.04, 5.05, 5.07–5.08 & 5.11
    - o Project: continue Phase 2
    - o Homework: begin Assignment 4; continue Assignment Group 3
- Week 9:
  - DSFS: Chapter 18
  - TDSH: Chapters 8.5.8 & 24.1–24.5
    - o Project: continue Phase 2
    - o Homework: continue Assignment Group 4; complete Assignment Group 3
- Week 10:
  - DSFS: Chapters 12, & 22 (as needed for projects)
    - o Project: complete Phase 2; presentations
    - o Homework: complete Assignment Group 4
- Week 11:
  - o Project: presentations

Notice: Appropriate Use of Course Materials

It is important to recognize that some or all of the course materials provided to you may be the intellectual property of Drexel University, the course instructor, or others. Use of this intellectual property is governed by Drexel University policies, including the policy found here: https://drexel.edu/it/about/policies/policies/01-Acceptable-Use/

Briefly, this policy states that course materials, including recordings, provided by the course instructor may not be copied, reproduced, distributed or re-posted. Doing so may be considered a breach of this policy and will be investigated and addressed as possible academic dishonesty, among other potential violations. Improper use of such materials may also constitute a violation of the University's Code of Conduct found here: https://drexel.edu/compliance-privacy-audit/compliance/policies/cpo-1-01/ and will be investigated as such.

Academic Policies

This course follows university, college, and department policies, including but not limited to:

- Academic Integrity: https://drexel.edu/provost/policies/academic-integrity/
- Academic Integrity Policy: https://drexel.edu/studentlife/community_standards/code-of-conduct/academic-integrity-policy/
- Academic Integrity Conduct Process: https://drexel.edu/studentlife/community_standards/code-of-conduct/academic-integrity-policy/academic-integrity-conduct-process/
- Disability Resources: https://drexel.edu/oed/disabilityResources/overview/
- Course Add/Drop Policy: https://drexel.edu/provost/policies/course-add-drop/
- Course Withdraw Policy: https://drexel.edu/provost/policies/course-withdrawal/
- Drexel Student Learning Priorities: https://drexel.edu/provost/assessment/outcomes/dslp/
- Support for Equality and Diversity: https://drexel.edu/oed/

Incomplete Policy

Incomplete grades are contingent upon instructor approval and will only be considered in extenuating circumstances beyond the student's control. The instructor is under no obligation to offer an incomplete grade. At least 80% of the graded coursework must have already been completed in order for an incomplete grade to be considered (per the recommendation of the Provost's Office). An incomplete contract with an instructor-determined due date for delivery of the completed work must be completed by the student and the instructor. It can be found here: http://www.drexel.edu/provost/policies/pdf/forms/incomplete.pdf

Class Cancellation

On rare occasions, instructors may be delayed or unable to attend a scheduled class due to unforeseen circumstances. In the event that an instructor does not appear in class and has not notified the class of his/her expected arrival time, class is cancelled 15 minutes after the

scheduled start of class. More information about class cancellations can be found at:
https://drexel.edu/provost/policies/cancellation_instructor_absence/

Course Change Policy

The instructor(s) may, at his/her/their discretion, change any part of the course before or during the term, including assignments, grade breakdowns, due dates, and schedule. Such changes will be communicated to students via the course web site. This web site should be checked regularly and frequently for such changes and announcements.

Student with Disability Statement

Students requesting accommodations due to a disability at Drexel University need to request a current Accommodations Verification Letter (AVL) in the ClockWork database before accommodations can be made. These requests are received by Disability Resources (DR), who then issues the AVL to the appropriate contacts. For additional information, visit the DR website at https://drexel.edu/oed/disabilityResources/overview/ , or contact DR for more information by phone at 215.895.1401, or by email at disability@drexel.edu.