DATA5000: Introduction to Data Science

Winter 2026
Institute for Data Science
Carleton University

Description

The course covers topics relevant to data science: working with data, exploratory data analysis, data mining, machine learning. The concepts are illustrated using the R language. Students also receive hands-on tutorials (e.g., IBM Cognos Analytics, Tableau). Students will be evaluated by their course projects. The course will be lecture-based and will also offer some hands-on tutorials. The project component will be flexible and will involve data collection, manipulation, and analysis.

The project forms an integral part of this course. The project is to be completed in group of two-three students. Each group would have one or two technical expert(s) (a student from Computer Science, Systems and Computer Engineering, Information Technology, Physics, Chemistry), and one or two domain expert(s) (e.g., from Communication, Geography, Biology, History, Psychology, Economics, Business, Health Sciences, Cognitive Science, Public Policy and Administration, International Affairs). Domain experts may contribute to finding the right problem, justifying why it is important to study it, extracting the value and implications of the work. Technical experts do the heavy lifting of building models. The main goal for students is to learn how to work on a multidisciplinary team, i.e., for domain experts, it is about learning technical terminology, while for technical experts, how to fruitfully work with domain experts.

Class Schedule

Lectures are held every **Tuesday from 11:35 AM to 2:25 PM via Zoom**.
There will be no classes on February 17 (Winter break) and March 31 (Data Day).
Discord server will be used for course communication, news, and reminders. Please join the server here: https://discord.gg/SJEHUjezXM.
Schedule of seminars is posted on the course website.
Winter break: February 16–20, 2026 (no class).

Course Website

http://olgabaysal.com/teaching/winter2026/data5000_w26.html

Instructors

Dr. Olga Baysal
Email: OlgaBaysal@cunet.carleton.ca
Office Hours: via Zoom by appointment only
Website: http://olgabaysal.com/

Dr. Ahmed El-Roby
Email: AhmedElRoby@cunet.carleton.ca
Office Hours: via Zoom by appointment only
Website: https://people.scs.carleton.ca/~ahmedelroby/

Dr. Elio Velazquez
Email: ElioVelazquez@cunet.carleton.ca
Office Hours: via Teams by appointment only

Learning Material(s) and Other Course/Lab-Related Resources

No textbooks are needed in the course.
Students are not required to purchase textbooks or other learning materials for this course.

Topics Covered and Learning Outcomes

- Foundations of Data Science and the Data Science Lifecycle
- Exploratory Data Analysis and Data Visualization
- Data Wrangling and Cleaning Techniques
- Introduction to Probability and Statistical Inference
- Supervised Learning: Regression and Classification
- Unsupervised Learning: Clustering and Dimensionality Reduction
- Working with Real-world Datasets (e.g., open data, health, transportation)
- Introduction to Natural Language Processing
- Reproducibility and Scientific Computing Practices

By the end of this course, students will be able to:

- Explain the key stages in the data science workflow and apply them to real-world data problems.
- Perform exploratory data analysis and generate meaningful visualizations using R or Python libraries.
- Clean and prepare raw datasets for analysis through appropriate data wrangling techniques.
- Apply fundamental statistical concepts to analyze data and interpret results.
- Implement basic machine learning models for classification, regression, and clustering tasks, and evaluate their performance.
- Use real-world datasets to draw insights and communicate findings effectively through visual, written, and oral presentations.
- Develop reproducible code and workflows using scientific computing tools.

Course Schedule

- Tuesday January 6 - Lecture 1: What is Data Science?/ Introduction to R.
- Tuesday January 13 - Lecture 2: Working with Data.
- Tuesday January 20 - Lecture 3: Visualization and Exploration.
- Tuesday January 27 - Lecture 4: Data Mining and Machine Learning I.
- Tuesday February 3 - Lecture 5: Machine Learning II.
- Tuesday February 10 - Paper presentations.

- Tuesday February 17 - No class (Winter break).
- Tuesday February 24 - Tutorial OR guest lectures.
- Tuesday March 3 - Tutorial OR guest lectures.
- Tuesday March 10 - Tutorial OR guest lectures.
- Tuesday March 17 - Tutorial OR guest lectures.
- Tuesday March 24 - Project presentations I.
- Tuesday March 31 - No class (everyone attends Data Day 12.0).
- Tuesday April 7 - Project presentations II.

Evaluation

- Project proposal: 15% (due January 25, 11:59 PM)
- Paper selection: 0% (due January 27, 11:59 PM)
- Paper presentation: 15% (February 10)
- Poster abstract submission: 5% (TBD)
- Project presentation: 15% (March 24 or April 7)
- Project report: 50% (due April 14, 11:59 PM)

University Policies

For information about Carleton's academic year, including registration and withdrawal dates, see Carleton's Academic Calendar.

Academic Accommodations

Carleton is committed to providing academic accessibility for all individuals. Please review the academic accommodation available to students here: https://students.carleton.ca/course-outline/.

Academic Integrity

Student Academic Integrity Policy. Every student should be familiar with the Carleton University Student Academic Integrity policy. A student found in violation of academic integrity standards may be sanctioned with penalties which range from a reprimand to receiving a grade of F in the course, or even being suspended or expelled from the University. Examples of punishable offences include plagiarism and unauthorized collaboration. Any such reported offences will be reviewed by the office of the Dean of Science.

More information on this policy may be found on the ODS Academic Integrity page: https://science.carleton.ca/students/academic-integrity/.

ChatGPT/Generative AI Usage

The use of any AI system will be considered academic misconduct. This includes, but is not limited to, chatbots or code generators (e.g., ChatGPT) for projects (reports, models, etc.). An exception to the above rule is made for automated grammar and punctuation checking tools (such as Grammarly).

Plagiarism

As defined by Senate, "plagiarism is presenting, whether intentional or not, the ideas, expression of ideas or work of others as one's own". Such reported offences will be reviewed by the office of the Dean of Science. More information and standard sanction guidelines can be found here: https://science.carleton.ca/students/academic-integrity/. Please note that content generated by an unauthorized A.I.-based tool **is** considered plagiarized material.

Unauthorized Collaboration

Senate policy states that "to ensure fairness and equity in assessment of term work, students shall not co-operate or collaborate in the completion of an academic assignment, in whole or in part, when the instructor has indicated that the assignment is to be completed on an individual basis".

Student Rights & Responsibilities

Students are expected to act responsibly and engage respectfully with other students and members of the Carleton and the broader community. See the 7 Rights and Responsibilities Policy for details regarding the expectations of non-academic behaviour of students. Those who participate with another student in the commission of an infraction of this Policy will also be held liable for their actions.

Student Concerns

If you have any concerns regarding this course, your first point of contact is me. Please email me or visit during my student hours, and I will do my best to address your concerns. If I cannot resolve the issue, the next point of contact is the Institute for Data Science at cuids@carleton.ca. If the concern remains unresolved, the final point of contact is the Office of the Dean of Science at ODScience@carleton.ca. Please follow this order of contact.
Note: You can also bring your concerns to Ombuds services.