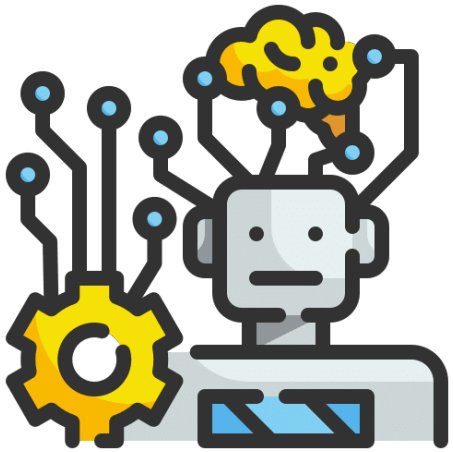
Data Analytics

With a focus on machine learning

Winter, 2025



Essential Information:

- **Professor:** Carl Toews (adjunct faculty, he/him)
- **Email:** Carl.Toews@evergreen.edu (give me a day or two to get this up and running)
- **Phone:** (253) 327-4659 (This is my personal cell; please text for emergencies, otherwise use email).
- **Office hours**: if you want to chat outside of our seminar, just shoot me an email and we can set up a Zoom call. I am currently working full time as a data scientist in Bellevue, so my availability 9-5 is fairly limited, but my evenings are pretty flexible.
- **Class hours:** Tuesdays, 6:00pm-9:00pm

Course Information:

Welcome to Data Analytics! The objective of this course is to develop skill in the art and science of extracting insights from data. A significant focus will be on machine learning, which is the process of using data, algorithms, and statistics to make predictions. Other elements of data analytics include data cleaning, data visualization, crafting reports and making recommendations, all of which you will have a chance to practice. The course will be taught from an applied perspective, i.e. it will concentrate on developing practical intuition and marketable skills. However, since a basic understanding of mathematical underpinnings can help safeguard against algorithmic misuse, we will also cover theory as needed.

Specific Learning Goals:

After successfully completing this course, you should be able to:

1. use Pandas to manipulate data efficiently
2. effectively use a wide subset of the Scikit-learn machine learning library
3. generate informative and appropriate data visualizations using Python plotting tools

4. choose appropriate machine learning models for diverse data analysis tasks
5. explain in detail how each type of machine learning model works
6. describe the essential steps of a full machine learning life cycle, and execute these steps effectively for a range of predictive tasks
7. produce well-structured data analysis reports that communicate data-centric insights to a specific target audience
8. understand the assumptions and relevant statistics behind an A/B test
9. understand how to fine-tune machine learning models to generate better performance

More broadly, this class should serve to improve your capacity to think critically about how data can be used to generate insights, make predictions, suggest actions, and generate value.

Required Text:
"Hands On Machine Learning with Scikit Learn and Tensorflow, 3rd Edition", by Aurélien Géron. (Other editions will be OK to get going, but there was a big leap between the 1st and 2nd editions, and a small one between the 2nd and 3rd).

Other Requirements:

- Laptop capable of running Python. (If you don't have one, talk to me).
- A lab notebook (marble covered, 9.75x7.50)

Assessment:

Participation	20%
Individual Homework	25%
Group Homework	25%
Final Project	15%
Final Exam	15%

Class Expectations:

This class will be taught in an "inquiry based" style, meaning that instead of allocating our class-time to lectures, we will use most of our time together to actively pose, debate, and answer questions that get at the heart of the material we are trying to master. Inquiry based learning (IBL) emerges from the philosophy that learning works best when the learner *discovers* ideas rather than simply has them presented in finished form: to really understand a subject, the learner needs to think not just about how to *solve* problems, but also about how to *pose, assess*, and *communicate* them. In describing an IBL classroom, a colleague of mine noted that "Classrooms shouldn't be places where young people come to watch old people work." Our classroom won't be. In this setting, my role as the professor is to help guide you towards productive discovery, and your job as the learner is to be curious, engaged, and communicative.

There are a number of ways to structure an IBL class, but salient features of this class will include the following:

- Hands-on activities instead of lecture

- An emphasis on group work
- Lots of dialogue and discussion
- Student presentations
- A supportive environment in which to take "risks"
- An emphasis on communication, both oral and written
- A need for trust and confidence, both student-student and student-professor

A typical day in my classroom will start with me posing one or two questions that pertain to material we have already covered. The problems will either be abstract thought experiments, or tie into the previous day's homework, and I'll call on students to provide answers. New material will be presented in summary form, generally a brief presentation, or possibly a guided computation. Much of the class will consist of an exploration of this new material in small groups, punctuated with student presentations of select solutions and, hopefully, some lively discussion, all held together with occasional professorial input. The questions we address in class won't always exactly mirror the problems in the book, but by working through them, and presenting solutions to your peers, you will develop the technical and communicational skills needed to solve both book problems and problems from the broader world.

Productive failure is an idea that lies at the root of our approach. When you're trying to learn something, never making a mistake is generally a sign that you're not pushing yourself hard enough. This class should be a safe and supportive space in which to get things wrong. When talking or presenting, you are challenged to work slightly outside of your comfort zone, to volunteer answers when you have a pretty good idea but aren't 100% certain, to risk a conjecture that might turn out to be off the mark. And when you are listening to fellow students talk, you are challenged to pay strict attention, to flag small errors of language or comprehension, and to politely and respectfully help guide one another towards a clearer and truer picture of the matter at hand. Failure is part of the design spec for this class, and it can be hard, but you will not be struggling alone.

Although the spirit of what I'm shooting for with this IBL style class is probably clear, here is a minimalist list of concrete expectations:

- attend all classes
- do all assigned homework
- participate actively in class discussions and class group work activities
- volunteer to present solutions
- volunteer answers to questions I pose, and to ask your own questions when you are confused, uncertain, or simply thinking outside the box
- be courteous and supportive of your fellow learners
- help create a classroom that is a supportive, energetic, respectful place to learn.

More broadly, my basic hope and expectation is that you will engage the spirit of Inquiry Based Learning with enthusiasm, openness, and joy (it *is* fun), and help make this class a fun and supportive place in which to learn statistics.

Lastly, a word about goals and outcomes: the goal in IBL is to produce critical thinkers who have a strong, creative command of the subject material. For me, one of the strongest elements of IBL is the host of secondary skills you develop almost "for free", including skills in abstract thinking, working from first principles, working with other people, and communicating your ideas. As a colleague of mine has noted, "All of the secondary skills you will develop in this course are highly

valued by society. Whether you become a teacher, a lawyer, an engineer, or an artist, what differentiates you from your competition is your ability to think critically at a high level, collaborate professionally, and communicate effectively."

Details About Class Activities:

Class rhythm:

Our class meets once a week on Tuesdays for three hours. Three hours is a lot of time. We're going to break it up in a few ways:

- We will often start off by discussing a question or two that I pose on the board. The question might be related to the previous week's reading, or to stuff that we intend to work on in class that day, or simply some abstraction that has direct or indirect bearing on the content of the class. This is a chance to chat in small groups, share your thoughts, stretch your brain.
- We will also spend some time discussing any questions that might have cropped up from the previous week's reading. Formulating such questions is part of your weekly individual homework.
- Most weeks you will be expected to submit some sort of data project, generally in a group of 3-4 people. Each week I will ask one or two groups to go over their solutions. (More on that below). If your role within your group is "presenter", you should expect to come to talk the class through your work, explain your conclusions, field questions, etc.
- We will plan to take a 10-15 minute break each day roughly half-way through the session. I'll try to bring some snacks to make it less tedious; others are invited to do the same.
- New material will be presented in a fairly compact format. I generally do not like using powerpoint, but because of the hybrid nature of the class, it might be the most efficient way forward. The goal of the "lecture" is to provide some framing for the material that you're about to tackle, and maybe a bit of guidance in reasoning through the gotchas.
- Time permitting, the seminar will wrap up by forming groups and starting to brainstorm some of the computational work for the following week. Groups will be expected to meet asynchronously at least once over the following week to finish this work up.

Group homework (computational):

In industry, the importance of being able to work in a group is hard to overstate. To mimic that dynamic, group work will be a critical part of what we do in this class. In small peer groups, you will be able to discuss and evaluate ideas, figure out how to formulate solutions, and ultimately present these solutions to the class at large. The peer group thus provides ideas, feedback, support, and a set of other learning models.

Each week you will be assigned some computational project. In the beginning, these projects will be designed to bring you up to speed in, say, Python, while later on they will assume the form of full-fledged data challenges. Expect to work on these collaboratively, with a set of peers that I assign. Within each group, there might be specific roles assigned to specific people, e.g. "coder", "editor", "presenter", etc. You should plan on playing a range of roles throughout the semester.

Individual Homework (conceptual):

Each week I will assign some individual work to do. In general, this work involves reading a portion of the text and answering a few questions. Some of those questions could involve

computations, but in general, they will be conceptual, since the group projects will cover the computational side of things. Individual work can be typed or written up and submitted via Canvas.

Your homework will be assessed on a 0-3 scale, as follows:

3	All problems attempted; correct answers, strong effort to articulate thought process
2	Most problems attempted; correct solutions, but not a great articulation of thought process
1	Half or fewer of the problems attempted; incorrect and ill-articulated answers
0	Nothing submitted

***Final Project*:**

One of the cool things about data science is that you can use it to make a significant impact on just about everything. One of your tasks this quarter is to find a question or concern that interests you, and use the techniques we have learned in this class to pose and answer questions about it. Each project will involve data collection, statistical analysis, and a formal, typed write-up. You will be expected to use your computer to process the data, form plots, and perform analysis, and your final write-up is expected to be a polished, professional looking report. The final project will be done as an individual, not as a group, and I will pass out an assessment rubric in advance.

***Final Exam**:*

The final exam will be a mock interview for a data scientist position focusing on machine learning. Depending on how constrained for time we are, this interview might be oral or written, but in any case, the best preparation for it is a thorough understanding of the concepts presented in the text.

Participation:

Since active participation by each member of the class is a critical element in the success of the IBL classroom, participation counts as a significant portion of your final assessment. Here is what I look for:

- *Engaged attendance*
  Be present in body and mind. The former is easy to check. The latter requires some judgment on my part, but I assess mental engagement via the questions you ask, your contributions to the class dialogue, your willingness to work in a team, etc.

- *Presentations of solutions*
  Everyone will get a chance to present their computational work at some point. Technical demos are hard! All I ask is that you get up there and do your best. Remember, the point of a presentation is not to convince me that you've done a problem, but rather make the problem and its solution clear to your peers. Presenters should explain their reasoning, not simply show answers. Fellow students can ask questions at any point, and should pay attention both to what is written on the board and what is said by the speaker.

Provisional Calendar:

The following is a rough example of how we might pace this class. I have left some wiggle room in the schedule deliberately to allow ourselves to be guided by how our interests evolve.

Week	Topic	Reading (due before class)	Computational work (started after class)
1: January 7	The data analysis landscape	–	EDA + Python fundamentals
2: January 14	The machine learning landscape + end-to-end ML project	Chapters 1 and 2	Regression
3: January 21	Classification	Chapter 3	MNIST
4: January 28	Training details	Chapter 4	Polynomial regression
5: February 4	Support vector machines	Chapter 5	California housing data
6: February 11	Decision Trees and ensemble models	Chapters 6 and 7	Zillow challenge
7: February 18	Dimensionality reduction	Chapter 8	Swiss roll
8: February 25	Tensorflow and neural nets	Chapter 9	TBD
9: March 4	TBD	TBD	TBD
10: March 11	TBD	TBD	TBD

Policies:

Attendance:

Built into the philosophy of inquiry-based learning is the idea that we help one another to learn. As a consequence, your daily attendance in class is very important, not just for your own benefit, but for that of your peers. I expect you to be present. Unexcused absences will be noted and factor into your participation grade.

Late work:

I am not a huge fan of late work, but life happens. In general, I expect work to be on time unless there are compelling reasons for it to be late. If you are having trouble making a deadline, let me know (in advance, ideally) and we can troubleshoot together.

Planned Absences:

If you need to be absent for some family or medical emergency, you should contact me in advance (if possible) or as soon as possible after the emergency.

Classroom policies:

You are welcome to bring a cup of coffee or a bottle of water to class, but please eat your meals outside of class (except for the snack we plan to have in the middle of each class). Please turn off your phones and keep your laptops closed, unless we happen to be doing a computer exercise. You can take a bathroom break if necessary, but please make this the exception, not the rule—in general, I don't want people entering and leaving the room during class.

Academic integrity:

This is a collaborative classroom. But the difference between your work and other people's work needs to be acknowledged. Please understand and abide by the university's academic integrity policy.

Disabilities:

If you have a physical, psychological, medical or learning disability that may impact your course work, please contact me. All information and documentation is confidential.

Getting Help:

My contact information is at the top of the syllabus. Please reach out! At the moment I have a day job that keeps me pretty busy between 9 and 5, but I will definitely respond to emails in a timely fashion, and I'm always happy to jump on a Zoom call to discuss anything.