

Data Science for Business : Technical

Spring 2025: GB 2336/UB 57

Professor: Dr. Chris Volinsky
Email: chris.volinsky@stern.nyu.edu
Office: KMC 8-81
Office Hours: TBD

Teaching Fellows: TBD

Class Time:

- **GB 2336** : Wednesdays 6-9PM
- **UB 57** : MW 9:30 – 10:45 / 11:00 – 12:15

Course Objectives:

This course will teach you how to think about data-centric problems in the business world through the lens of data science. We will cover the fundamentals of the data science process and how predictive models and machine learning are used to find solutions for business problems using real-world data.

We will focus on data-analytic thinking: how to approach problems, how to develop insights using data, how to apply machine learning algorithms, how to validate and assess algorithmic results, and how to communicate those results to a business audience appropriately. This is an applied class, with analysis of real data sets using Python's powerful data science libraries (pandas, sklearn, matplotlib, and others)

Throughout the semester we will examine examples of data science in practice through class discussions of real-world applications of data science.

After taking this course you should:

1. *Approach business problems data-analytically.* Think carefully & systematically about whether & how data can improve business outcomes, to make better-informed decisions for product management, marketing, sales, operations, consulting, etc.
2. *Understand the fundamental principles of data science, including the entire data science process from problem formulation to deployment.* Know when and how to apply different methods, and how to evaluate and communicate them.
3. *Have had hands-on experience implementing machine learning algorithms in Python.* This includes a group term project using data of your choosing. You will formulate a relevant business problem, and do a complete analysis, write-up and presentation.

Topics include but are not restricted to:

- Understanding the Data Science Process

- Exploratory Data Analysis
- Decision Trees
- Evaluating Predictive Models
- Linear Classification and Regression
- Variable Selection and Complexity Control
- Modelling Text Data
- Unsupervised Models and Clustering
- Recommender Systems
- Ensemble Methods
- Neural Nets, Deep Learning, and AI
- Data Science Ethics

Prerequisites:

Python: It is required that you have some programming experience in Python prior to taking this class. You do NOT have to be a Python expert or a programming whiz, however, basic proficiency with Python is expected. Any Introduction to Python class and some experience with programming should suffice. I will **not** be teaching Python basics or fundamentals in this class. You are expected to come into class understanding basic Python concepts such as: How to use a notebook environment (e.g. Jupyter), how to link and use different libraries (e.g. numpy and pandas), different data types, classes and structures, how to input data, and how to write and use simple functions.

Statistics: At least one introductory Statistics, data analysis, machine learning, or data science class is required. Knowledge of basic statistics – measures of central tendency and variation, basic laws of probability and statistical concepts like confidence intervals, statistical significance and hypothesis testing will be assumed.

Textbook:

The **required** textbook for the class will be the following:

[DSB] *Data Science for Business*. By Provost & Fawcett (O'Reilly, 2013) – 2nd Edition

The book is available as a coursepack through Brightspace. In order to have access to the book, you must purchase a digital rights key code from the bookstore. Details will be provided at the start of class.

This book covers the fundamental material that will provide the basis for you to think and communicate about data science and business analytics.

There are **optional** secondary books I will refer to during the semester that will be helpful in getting a fully rounded view of the material. They are available for free:

[ISL] *An Introduction to Statistical Learning (with Applications in Python)*, by James, Witten, Hastie, Tibshirani, Taylor. Available free at <http://statlearning.com>. [Here is the link](#) to download the Python version of the book. This book is a bit technical for students who want to go deeper, and provides more statistical background than we will cover in class. It also goes into other algorithms and methods that we will not cover in class, but you might want to consider for your class project. Additionally the book contains detailed Python snippets to apply these methods to real data.

[Shmueli] *Data Mining for Business Analytics (Applications in Python)*, by Shmueli, Bruce, Gedeck, and Patel. Available freely in digital form at NYU Libraries. Covers similar material to what we are covering with many excellent Python examples.

Other Useful Resources:

[Pandas] *Pandas Workout: 200 Exercises to make you a stronger data scientist*. Excellent repository of examples for those who want to sharpen Python coding skills on interesting problems. Available free online via NYU Libraries.

[OpenIntro] *OpenIntro Statistics*. PDF available for free at <https://www.openintro.org/book/os/>. An online introductory statistics course. Excellent for brushing up on statistical topics and data-oriented thinking.

Software:

As noted above, some working knowledge of Python is required as a prerequisite. We will spend our time exploring the data analysis and machine learning libraries that you will use through the class (mainly pandas, sklearn, and matplotlib). I will provide several iPython notebooks that we will work through together and can serve as references to help you learn the basics of the key data science and machine learning libraries and functions. These notebooks will help you learn the tools you will need for your term project.

Class examples will use [Google's Colab platform](#). You must have access to a computer on which you can access and run the class Colab notebooks. If you do not have such a computer, please see me immediately so we can make alternative arrangements. You should bring your computer to class. If you need additional help with Colab or Python, please see your teaching assistant. Feel free to code using any Python IDE you are familiar with (VSCode, PyCharm, etc), but you will need to submit notebooks that we can run on Colab.

Statement on Generative AI:

Generative AI tools such as ChatGPT, Gemini, and Copilot have become a valuable part of the learning experience, and an important skill in the business world today. I encourage the use of these and other AI tools to aid in coding assistance, idea generation, and overcoming obstacles in your work. Specifically for coding, these tools can greatly help the speed of getting things done as well as the

speed of learning new concepts. In class I will demonstrate how to use these tools to aid in data science problem solving, coding, and understanding.

Please use these powerful tools wisely! The best learning approach is to attempt problems yourself, and reach for the power tools when you hit a stumbling block. In written examples like homeworks and your data project, use GenAI output as a suggestion and a guide to frame your own writing. **Note that direct copying and pasting content generated by AI without appropriate modification, critical thinking, or attribution can be considered plagiarism and violate Stern's academic integrity standards.**

Grading:

- Homework (Analytic Exercises/Reflections): 30%
- Quiz (2-3 in-class): 30%
- Term Data Project: 35%
- Participation / Attendance : 5% (Attendance will be taken at every class)

Grading Policies:

NYU Stern policy for core classes is that **no more than 35% of the class will receive a grade of A or A-**. Therefore, the point value cutoff for an A/A- grade will be scaled based on the performance of the class. Outside of that cap, the majority of students will receive B range grades, with 5% or less of the class receiving C+ or lower.

Although you will get your grades back from homeworks and quizzes quickly, you may not know how this transfers to a letter grade. Mid-semester, I will provide an assessment of where you are letter-wise so it will provide an opportunity for course correction if needed.

Homework: Homeworks are intended for you to practice your analytic and coding skills. The problems will be provided as iPython notebooks - you should copy and save to your own Google Drive, edit with your solutions, and submit into Brightspace as both .ipynb notebooks and .pdf files. Please add commenting to your code so we can see your reasoning in case the answer is wrong so we can give partial credit. For those not experienced with Python, the HW assignments can be substantial and may take significant time. Please start the assignments early and leave time to ask questions during class or during office hours so there is not a last minute crunch.

Homeworks will have a strict due date and time as noted in the class schedule. Homeworks submitted up to 24 hours late will have a 25% penalty, and up to three days late will be a 50% penalty. Homeworks can be done collaboratively, but each student is required to submit their own unique homework and is responsible for their own learning.

Readings and Articles: Please note the readings and articles that are assigned and do the reading before the associated class. Questions may come up in class, and you may be called on randomly. Reading material is fair game for exams.

Quiz: In-class exams will cover material from lectures and readings – and will **not** have a coding requirement. They are closed book but you will be allowed one single page of notes front and back. **Make-up quizzes will not be offered** – if you have a hardship requiring you to miss a quiz, you will need to let me know with at least a week’s notice, with a documented reason.

Participation/Attendance: This is a measure of a student’s engagement in the course via 1) Attendance and 2) Class participation during in-class and online discussions. It will be graded on the scale of 1=minimum engagement, 3=expected engagement and 5= exceptional engagement. I will take attendance via a sign in sheet at the beginning of each class. Please let me know if you will not be attending class – each student may have up to 2 absences without impact on this grade. *Having someone else sign your name to the attendance sheet is considered a violation of Stern Code of Conduct and will be subject to appropriate disciplinary action.* Additionally, late arrivals to class and early departures will be noted and may affect this aspect of your grade.

I will attempt to get to know every student by name, but it is a challenge! Helping me learn who you are is a good way to do well on this grade. **Please use your name placards regularly throughout the semester** so that I can best learn your name.

Note: All classes will be video recorded and posted on the Content => Mediasite tab in Brightspace.

Term Data Project: Your class project will consist of: defining a business problem that can be addressed by data; writing up and proposing the analysis; collecting the data; cleaning and preparing the data; performing the proper analysis; providing an appropriate evaluation and defense of the model selected; and communication of the results via a presentation. You will be working in groups of 4 to do the project – you can pick the groups if you like, otherwise they will be assigned for you.

Learning how to work within a team is one main point of the project. Some of you will be better coders, others will be better communicators, and the team should make use of the skills of all members team in the best way possible while trying to balance the responsibilities. You will be asked to evaluate the contributions of your team members at the end of the project.

You will have several benchmarks over the course of the term to meet in order to successfully complete the project, including an initial proposal writeup of the business question, and a “plan of attack”, and a project update around the middle of the term. Your team will be expected to provide a writeup of the project and present the project and solution in a class presentation the last week of the class. More information will be provided.

Other Policies:

- **Religious Holidays:** NYU’s [Calendar Policy on Religious Holidays](#) states that members of any religious group may, without penalty, absent themselves from classes when required in

compliance with their religious obligations. **You must notify me at the start of the semester** of religious holidays or observances that might coincide with exams, assignments, or class times to schedule mutually acceptable alternatives. Please review all class dates at the start of the semester and review all course requirements to identify any foreseeable conflicts with exams, course assignments, projects, or other items required for participation and attendance.

- **Accommodations:** If you will require academic accommodation of any kind during this course, you must apply through the [Moses Center for Student Accessibility](mailto:mosescsa@nyu.edu) (212-998-4980, mosescsa@nyu.edu,). The Moses Center will send me an official notice of accommodation. If you will need to take an exam at the Moses Center for Student Accessibility, you must submit a completed Exam Accommodations Form to them at least one week prior to the scheduled exam time to be guaranteed accommodation.
- **In-class:** You should bring a laptop to class with you if possible, as it will be helpful for you to be working through the Python notebooks with me in real time. If you do not have a laptop, I can make sure you are in a group that does have one. You may use a laptop/tablet to take notes. *Please refrain from the use of mobile phones during class.*
- **Inclusion:** This course strives to support and cultivate diversity of thought, perspectives, and experiences. I am committed to every student feeling a sense of belonging in the classroom. If for any reason you feel excluded, I ask that I be your first point of contact to discuss your concern. I am open to hearing your concerns and learning from each other.
- **Academic Integrity:** We take pride in our well-rounded education and approach our academics with honesty and integrity. Indeed, integrity is critical to all that we do here at NYU Stern. As members of our community, all students agree to abide by the [NYU Academic Integrity Policies](#) as well as the NYU Stern Student Code of Conduct, which includes a commitment to:
 - Exercise integrity in all aspects of one's academic work including, but not limited to, the preparation and completion of exams, papers and all other course requirements by not engaging in any method or means that provides an unfair advantage.
 - Clearly acknowledge the work and efforts of others when submitting written work as one's own. Ideas, data, direct quotations (which should be designated with quotation marks), paraphrasing, creative expression, or any other incorporation of the work of others should be fully referenced.
 - Refrain from behaving in ways that knowingly support, assist, or in any way attempt to enable another person to engage in any violation of the Code of Conduct. Our support also includes reporting any observed violations of this Code of Conduct or other School and University policies that are deemed to adversely affect the NYU Stern community.
- **Stern Code of Conduct:** The Stern Code of Conduct and Judiciary Process applies to all students enrolled in Stern courses. Information can be found here: <https://www.stern.nyu.edu/uc/codeofconduct>. The full Student Conduct Policy can be found here: <https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/university-student-conduct-policy.html>