

## BMKT 670 Fall 2020 Applied Data Analytics

### Welcome!

This course lays a technical foundation for work as an analyst or data scientist. The course has three main focuses:

- **Data Management:** How do we store data for easy retrieval? How do you clean data to prepare it for analysis?
- **Data Acquisition:** How do you use APIs and web scraping to build data sets for yourself?
- **Data Analysis:** There are more potential ways to analyze data than we can cover in a semester (or a degree, or a career!). We'll hit some key analytical techniques quickly, creating a body of knowledge that can be extended during your capstone, in other classes, or on the job.

Our course builds real world experiences and we will create an environment that will help you practice everyday skills to begin working in data science and analytics.

In Applied Data Analytics, you'll consume new information outside of class in the form of readings, recorded lectures, and other supplementary materials. The in-class time will be much more like a lab class in the hard sciences—an opportunity for you to do work and receive help quickly. Come to class prepared and ready to work hard for two hours. This class will have an explicit lab component: we will meet online on Thursdays from 4:00 to 6:00 to work on assignments, answer questions, and get help. For those of you who cannot attend these Thursday sessions live, please make sure to schedule time as needed throughout the semester.

We are covering a ton of ground here. This course is challenging, particularly for students with less technical experience. Please avail yourself of office hours and opportunities for extra help.

This syllabus has the following sections:

- **Meeting and Communication** How we'll work together this semester.
- **Course Objectives** A description of what I'm hoping you'll get out of this course.
- **Required Materials** The books required and recommended for the course.
- **Grading** An introduction to our contract grading this semester.
- **The Work** This section describes all the work that you could do in this course. You'll need to read this carefully before deciding on the work you'll contract for.
- **Class Format** A brief introduction to what we'll actually be *doing* in our class meetings.
- **Course Outline** An outline for what we'll be covering when during the course. Subject to change, of course.
- **Code of Contact** My expectations for you regarding things like behavior and collaboration.

### Meeting and Communication

Some key things to know about how we'll meet and how we'll communicate with each other:

- **Weekly Meeting:** our class will meet weekly on Mondays from 4-6. We'll work to avoid penalizing students who cannot meet at this time. Expect to meet on Zoom for those of you who can join synchronously: <https://umontana.zoom.us/j/8480321333>.
- **Weekly Lab:** We'll meet on Thursdays from 4-6 to get help on assignments or exercises from class. This lab is also available to the students in Text Mining.
- **Weekly Videos:** Each week I'll post a short(ish) video telling you what's going on this coming week, what you should be thinking about, any deadlines, etc. Pleasepleaseplease watch these videos.
- **Office Hours:** I have office hours at the time indicated above, but I'm happy to meet at other times if those times don't work for you.
- **Moodle:** All file sharing will take place via Moodle<sup>1</sup>.
- **Slack:** I like to use Slack as a collaboration software. Please use this [link](#) to join our UMT-Analytics team Slack. I strongly recommend installing the phone app and the desktop app. This is often the easiest way to collaborate, ask questions, and share information. Once you join the team, please join the channel #tswd. Let me know if you have any trouble.
- **Email:** I'll also use email occasionally to make announcements when I need to make sure they reach the whole class. I'll try to also post this stuff to Slack. Reach me on email via [john.chandler@business.umt.edu](mailto:john.chandler@business.umt.edu).
- **Text:** Having an emergency? Text me at 406-544-8720 any time you want. If I'm asleep I'll have notifications off. If you just call me without texting first, I'm unlikely to answer. Who answers unknown numbers these days?

## Course Objectives

1. Students will learn how to manage a data set and prepare it for analysis.
2. Students will learn how to use programming to acquire data sets.
3. Students will learn to embed analytical insights in a business context.
4. Students will receive additional practice and instruction with data science technology including Python, R, and SQL.
5. Students will be conversant in the key themes in the ethics of data science.

## Required Materials

1. *Weapons of Math Destruction*, Cathy O'Neil, Crown.
2. *Everybody Lies*, Seth Stephen-Davidowitz, Dey Street Books.
3. (Optional) *Data Science from Scratch*, Joel Grus, O'Reilly.
4. (Optional) *Algorithms for Data Science*, Steele, Chandler, and Reddy, Springer.
5. (Optional) *Data Science for Business*, Provost and Fawcett, O'Reilly.
6. (Optional) *The Statistical Sleuth*, Ramsey and Schafer, Brooks/Cole Cengage Learning, 2013.

## Grading in the Course

This course uses a type of grading you may not be familiar with, called "contract grading". If you haven't used it before, it may take some getting used to. **This course does not use points or letter grades**, other than the final grade you decide on (or "contract for") at the beginning and receive at the end. Instead of grades, you'll receive feedback on your work and guidance on bringing it up to the standards of the

---

<sup>1</sup> I think this is true. If it turns out not to be true, I'll mention it several times and let's all work to amplify the message.

course. Once your finished product is approved, the work will be considered completed and you can move on. Another uncommon aspect of this course: you're going to pick your own deadlines (within boundaries). I've explained a bit about *why* I like contract grading in a video posted to Moodle. The following list details the requirements for each grade. Each item is described more fully further down. In the first week of the course you'll choose your workload, your deadlines, and your desired grade.

### **Passing**

To pass the class, I expect the following from you:

- Honoring deadlines you're contracted for.
- Completing any reading for the week before Monday of that week.
- Participating fully in your topic discussion group.
- Communicating in a timely fashion with me if you are experiencing any issues with the course.
- Communicating in a timely fashion with your fellow students when you need to coordinate with them.
- Treating everyone involved in the class with respect.

### **Grade: C**

The C level is the path of least resistance. If you complete this level, you'll be exposed to the key ideas of the course, but you'll only do the minimum amount of work. You should choose this option if this semester is shaping up to be a difficult one for you and you don't need a grade higher than a C<sup>2</sup>.

- Everything from "Passing".
- *Weapons of Math Destruction*.
- No more than three absences from your discussion group.
- Completion of the "Wedge Data Engineering" assignment.
- Two "Beautiful Mistakes" posts.
- Three class exercise submissions.

### **Grade: B**

This is the standard level in the class. This grade represents a balance between workload and comprehensiveness. If you complete this level, you'll gain deeper experience with the key ideas of the course.

- Everything from "Passing" and C.
- *Everybody Lies*.
- One "Acquire and Analyze" project.
- Co-lead of discussion group and no more than two absences from your group.
- Four "Beautiful Mistakes" posts.
- Five class exercise submissions.

### **Grade: A**

---

<sup>2</sup> Note: to remain a student in good standing in the CoB graduate programs you need a minimum GPA of 3.0 (B average).

This grade represents a more substantial investment in the materials of the class. This level represents the everything you would typically have the option to do in a full-semester class with traditional grading. Here, however, you must do everything to a high standard.

- Everything from “Passing”, C, and B.
- Two “Acquire and Analyze” projects.
- Co-lead discussion group and no more than one absence from your group.
- Six “Beautiful Mistakes” posts.
- Eight class exercise submissions.

The following table has a summary of the work required by grade.

Item	Grade		
	A	B	C
<b>Books</b>			
"Weapons of Math Destruction"	1	1	1
"Everybody Lies"	1	1	0
<b>Wedge Data Engineering</b>			
Project & Write-up	1	1	1
<b>Acquire &amp; Analyze</b>			
Project & Write-up	2	1	0
<b>Beautiful Mistakes</b>	6	4	2
<b>Class Exercises</b>	8	5	3
<b>Discussion Group</b>	Miss 1	Miss 2	Miss 3

## The Work

There is a lot of work mentioned in the previous section. This section tells you more about what that work is.

### Books

There are two books assigned for this semester. The first, *Weapons of Math Destruction*, will be read by everyone. We’ll discuss this in class and write a short reflection paper on it. The second book, *Everybody Lies*, will only be read by students going for an A or B. Again, we’ll discuss the book and the assignment(s) for this book will be determined mid-semester.

Both books are informative and fun to read<sup>3</sup>. You will select your due dates in your contract form.

### Discussion Groups

In the first week you’ll be assigned to discussion groups. The purpose of these groups is to give you an opportunity to work with others in the class and build connections. This is particularly important as many students may be asynchronous during parts of the semester. Your discussion group will meet once a week for ten weeks. The meetings should be about 30 minutes (though I’d encourage you to schedule an hour). Each week a group leader will circulate two articles or blog posts. Each article will discuss a

---

<sup>3</sup> Your mileage may vary, but I thought they were fun.

technique or technology that we've covered in the class<sup>4</sup>. Want to read about something but not sure how to find a good article? Ask me! For the second article, if you can find an application of the technology or technique, that'd be great.

The group will read the articles and discuss. Considering how busy we all are, you might want to do the reading at the beginning of the meeting, [like they do at Amazon](#). Read the articles, discuss them, and share a summary of what you've discussed in the [Google Form](#) in the "Discussion Groups" section of Moodle.

During the first week of the semester I'll ask you to create a tentative schedule for your group meetings and indicate who the leaders will be. All students contracting for an A or B will evenly share the leadership role discussed above.

## Class Exercise Submissions

Many weeks we'll do work in class that supports the material covered in lectures or readings. Ten weeks will have code that you can optionally submit to have your work checked and to fulfill the requirements of your contract. Typically, these can be finished in 2-4 hours. You will submit your code for review and I'll ask you for revisions as needed to complete the work and format your code professionally.

Code will be submitted via an assignment in GitHub Classroom<sup>5</sup>

One sign of mature code is conforming to a style guide. When you work for a company, you'll probably have a style guide you use. Since we don't, I recommend the following guides for your work. Work for contract credit won't be accepted with major style guide violations.

- Python: Use the [Google Style Guide](#)
- R: Use either the [Google Style Guide](#) or the [Hadley Wickham Style Guide](#)
- SQL: I don't have a canonical style guide for SQL, but [this one](#) is quite good.

## Beautiful Mistakes

I'm not sure where the term "beautiful mistakes" comes from, although I know Bob Ross liked to use the term "[happy accidents](#)" and I watched a lot of him on PBS as a kid. My real inspiration is a lyric from R.E.M.'s song "[World Leader Pretend](#)" off the album *Green*:

*This is my mistake, let me make it good.  
I raised the wall and I will be the one to knock it down.*

Making mistakes is part of learning. If you're not making mistakes in this class, you're either extremely good at everything we're doing or not pushing yourself hard enough. I want to create a culture where we make mistakes, learn from them, and get better collectively. I'm going to ask you to document some of your programming mistakes in a Slack channel.

I've created a channel in our Slack called #beautiful-mistakes. During the semester I've asked you to post to that channel a variable number of times, depending on your contract. A post should look something like this:

---

<sup>4</sup> For instance, we begin with SQL, so an article on [SQL tricks](#) or [how it works under the hood](#) might be good choices.

<sup>5</sup> This is subject to change. I'm trying to bend GitHub Classroom to my will, with limited success so far.

- What you were trying to do.
- Your original code.
- A *clear explanation* of what was wrong with the original code.
- The corrected code.

These can be simple; there are many common mistakes that will be hugely beneficial. I'll occasionally quiz the class on the mistakes, so make sure to read the channel!

## Wedge Data Engineering Project

One of your “big” assignments will be to do a data engineering project based on data from the Wedge Co-op, the largest co-op grocery store in North America. This is a pretty big project that requires you to use Python to clean, process, and store 15 GB of grocery store receipt-level data. This project will dominate the first half of the semester and is required for all contract grades.

## Acquire and Analyze Project

This assignment will require you to find a data set and analyze it, though we will have a liberal interpretation of “analyze”. The data set does not need to be assembled programmatically, though that is encouraged. For the analysis you'll apply a technique we cover in class. The deliverable for this assignment will be your code and a write-up of your project.

## Class Format

Class format will primarily be hands-on work. Lectures will be delivered asynchronously via posted YouTube videos. The lectures will introduce new technical material, analyze real-world implementations of data science techniques, and serve as refreshers for the advanced marketing and technical material. You will also receive code to run and modify before our classes. The hands-on work in class will extend that work, so it is critical that you seek help if you cannot get the pre-work to run on your machine. Plan on sharing your screen during class, so make sure to mute any notifications that could be embarrassing.

Classes will be recorded and Zoom links will be posted as soon as possible after the class. (Typically processing the video file takes about 30 minutes.) Asynchronous students are encouraged to “skim” the class video. I'll often start classes with some comments on the lecture and reading. Then we'll collectively discuss the topics and the work for the day. Typically, there will be some boring parts where people are working—those should be skippable. Please let me know how I can make it easier for you to consume content asynchronously.

Our class meeting on Thursday, from 4:00-6:00, will be pure lab. I'm happy to discuss topics from readings or lecture, but this lab will co-convene with students from the Text Mining class. This is a time to make progress on your work with the ability to get help.

## Course Outline

The following is a rough outline of the topics to be covered, by week.

	Theme	Languages	Topics
<b>Week 1</b>	Data Management	Learning Python	Data Science Introduction, Python Overview, SQL Introduction
<b>Week 2</b>	Data Management	SQL	SQL Review
<b>Week 3</b>	Data Management	Google Big Query & SQL	GBQ, SQL

<b>Week 4</b>	Data Management	Python	APIs, Python & SQL
<b>Week 5</b>	Data Acquisition	Python	GBQ & Python
<b>Week 6</b>	Data Acquisition	Python	Wedge Work
<b>Week 7</b>	Data Acquisition	Python	APIs Part 2
<b>Week 8</b>	Data Management & Analysis	Python, R	APIs, <code>dplyr</code>
<b>Week 9</b>	Data Analysis	R	<code>ggplot</code>
<b>Week 10</b>	Data Analysis	R	Regression, Logistic Regression
<b>Week 11</b>	Data Analysis	R	ANOVA, $\chi^2$ tests
<b>Week 12</b>	Data Analysis	R	Tree-Based Models
<b>Week 13</b>	Data Analysis	R, Python	Shiny Apps, Agent-based Models, Clustering

## Code of Conduct

We are dedicated to providing a welcoming and supportive environment for all people, regardless of background or identity. We recognize that some groups in our community, however, are subject to historical and ongoing discrimination, and may be vulnerable or disadvantaged. Membership in such a specific group can be on the basis of characteristics such as gender, sexual orientation, disability, physical appearance, body size, race, nationality, sex, color, ethnic or social origin, pregnancy, citizenship, familial status, veteran status, genetic information, religion or belief, political or any other opinion, membership of a national minority, property, birth, age, or choice of text editor. We do not tolerate harassment of participants on the basis of these categories, or for any other reason.

Harassment is any form of behavior intended to exclude, intimidate, or cause discomfort. Because we are a diverse community, we may have different ways of communicating and of understanding the intent behind actions. Therefore, we have chosen to prohibit certain norms of behavior in our community, regardless of intent. Prohibited harassing behavior includes but is not limited to:

- written or verbal comments which have the effect of excluding people on the basis of membership of a specific group listed above;
- causing someone to fear for their safety, such as through stalking, following, or intimidation;
- the display of sexual or violent images;
- unwelcome sexual attention;
- non-consensual or unwelcome physical contact;
- sustained disruption of talks, events or communications;
- incitement to violence, suicide, or self-harm;
- continuing to initiate interaction (including photography or recording) with someone after being asked to stop; and
- publication of private communication without consent.

Behavior not explicitly mentioned above may still constitute harassment. The list above should not be taken as exhaustive but rather as a guide to make it easier to enrich all of us and the communities in which we participate. All interactions should be professional regardless of location: harassment is prohibited whether it occurs on or offline, and the same standards apply to both.

Enforcement of the Code of Conduct will be respectful and not include any harassing behaviors. Thank you for helping make this a welcoming, friendly community for all.

*This code of conduct is a modified version of that used by PyCon, which in turn is forked from a template written by the Ada Initiative and hosted on the Geek Feminism Wiki. This specific code of conduct can be found here: Greg Wilson (ed.): How to Teach Programming (And Other Things). Second edition, Lulu.com, 2017, 978-1-365-98428-0, <http://thirdbit.com/teaching>.*

## Names and Pronouns

Many people might go by a name in daily life that is different from their legal name. In this classroom, we seek to refer to people by the names that they go by. Pronouns can be a way to affirm someone's gender identity, but they can also be unrelated to a person's identity. They are simply a public way in which people are referred to in place of their name (e.g. "he" or "she" or "they" or "ze" or something else). In this classroom, you are invited (if you want to) to share what pronouns you go by, and we seek to refer to people using the pronouns that they share. The pronouns someone indicates are not necessarily indicative of their gender identity. This statement was found at [trans.umd.edu](http://trans.umd.edu) and you can visit that site to learn more.

## Double Dipping

A note on double dipping, which we define as submitting an assignment from one course in a second course. Here's what a recent syllabus for BMKT 680 says on the topic:

Please note that it is a form of academic misconduct to submit work that was also used in another course, aka "double dipping." **Don't do it.** If you are trying to get synergies across your classes/assignments, just ask a professor for advice. Don't try for a two-fer without approval!

I'm generally okay with double dipping if you get my approval, but I include the above quote to highlight that my stance is anomalous. If you're interested in using a project in my class for another class, let's talk about it and decide how you'll differentiate the two bodies of work. We *expect* you to use work from ADA in your capstone and don't consider that double dipping. You cannot submit an Acquire & Analyze project as-is for a Text Mining assignment.

## Additional "fine print"

**Professional Business Conduct in Class:** You are preparing to enter the business world as professionals and to prepare for a business career, so I expect each of you to behave in a professional manner in class.

- Arrive on time and stay for the entire class (unless excused by me).
- Behave with honesty and integrity. Don't let your team down!
- Respect everyone in class and listen openly to their ideas.
- Come to class prepared for discussion.
- Refrain from engaging in behavior that disrupts the class- this means no cell phones!

If at any time you are displaying disrespectful behavior, you may be asked to leave.

**Academic Integrity:** Academic misconduct is any activity that may compromise the academic integrity of the University of Montana. Academic misconduct includes, but is not limited to, deceptive acts such as cheating and plagiarism. Please note that it is a form of academic misconduct to submit work that was previously used in another course.

"Plagiarism is the representing of another's work as one's own. It is a particularly intolerable offense in the academic community and is strictly forbidden. Students who plagiarize may fail the course and be remanded to the Academic Court for possible suspension or expulsion."

"Students must always be very careful to acknowledge any kind of borrowing that is included in their work. This means not only borrowed words *but also ideas*. Acknowledgement of whatever is not one's own original work is the proper and honest use of sources. Failure to acknowledge whatever is not one's own work is plagiarism." So, ALWAYS err on the side of caution by citing the resources used in preparing your work. Moreover, always use direct quotations for exact wording taken from another source.

All students must practice academic honesty. Academic misconduct is subject to an academic penalty by the course instructor and/or a disciplinary sanction by the University. All students need to be familiar with the Student Conduct Code. The Code is available for review online at [http://life.umt.edu/vpsa/student\\_conduct.php](http://life.umt.edu/vpsa/student_conduct.php). It is the student's responsibility to be familiar the Student Conduct Code.

**Basic Needs Security** Any student who faces challenges securing food or housing, and believes that this could affect their performance in this course, is urged to contact any or all of the following campuses resources:

1. **Food Pantry Program:** UM offers a food pantry that students can access for emergency food. The pantry is open on Tuesdays from 9 to 2, on Fridays from 10-5. The pantry is located in UC 119 (in the former ASUM Childcare offices). Pantry staff operate several satellite food cupboards on campus (including one at Missoula College). For more information about this program, email [umpantry@mso.umt.edu](mailto:umpantry@mso.umt.edu), visit the pantry's website (<https://www.umt.edu/uc/food-pantry/default.php>) or contact the pantry on social media (@pantryUm on twitter, @UMPantry on Facebook, um\_pantry on Instagram).
2. **ASUM Renter Center:** The Renter Center has compiled a list of resources for UM students at risk of homelessness or food insecurity here: <http://www.umt.edu/asum/agencies/renter-center/default.php> and here: <https://medium.com/griz-renter-blog>. Students can schedule an appointment with Renter Center staff to discuss their situation and receive information, support, and referrals.
3. **TRiO Student Support Services:** TRiO serves UM students who are low-income, first-generation college students, or have documented disabilities. TRiO services include a textbook loan program, scholarships and financial aid help, academic advising, coaching, and tutoring. Students can check their eligibility for TRiO services online here: <http://www.umt.edu/trioss/apply.php#Eligibility>.

Please contact me any time for help if you are comfortable doing so. I will do my best to help connect you with additional resources.

**Disability Accommodations:** Students with disabilities will receive reasonable accommodations in this course. To request course modifications, please contact me within the first two weeks of class. I will work with you and Disability Services in the accommodation process. For more information, visit the Disability Services website at <http://www.umt.edu/dss/> or call 406.243.2243 (Voice/Text).

## COLLEGE OF BUSINESS MISSION STATEMENT

The University of Montana's College of Business is a collegial learning community dedicated to the teaching, exploration, and application of the knowledge and skills necessary to succeed in a competitive marketplace.

**Email:** According to University policy, faculty may only communicate with students regarding academic issues via official UM email accounts. Accordingly, students must use their GrizMail accounts ([netid@grizmail.umt.edu](mailto:netid@grizmail.umt.edu) or [fname.lname@umontana.edu](mailto:fname.lname@umontana.edu)). Email from non-UM accounts will likely be flagged as spam and deleted without further response. To avoid violating the Family Educational Rights and Privacy Act, confidential information (including grades and course performance) will not be discussed via phone or email.

## COLLEGE OF BUSINESS- ASSESSMENT AND ASSURANCE OF LEARNING

As part of our assessment process and assurance-of-learning standards, the School of Business Administration has adopted seven learning goals for our undergraduate students:

- Learning Goal 1 – CoB graduates will possess fundamental business knowledge.
- Learning Goal 2 – CoB graduates will be able to integrate business knowledge.
- Learning Goal 3 – CoB graduates will be effective communicators.
- Learning Goal 4 – CoB graduates will possess problem solving skills.
- Learning Goal 5 – CoB graduates will have an ethical awareness.
- Learning Goal 6 – CoB graduates will be proficient users of technology.
- Learning Goal 7 – CoB graduates will understand the global business environment in which they operate.

### Upon successful completion of this course, a student will be able to:

- Understand the overall lifecycle of a data science project and apply this knowledge to creation and communication of a project plan.
- Formulate a data science question from a business question. This process includes identifying metrics and data sets used to answer the question.
- Understand the challenges in assembling data sets at a variety of scales and will be able to form a well-reasoned opinion on how to approach the problem.
- Prepare a data set for analysis. This Extract-Transform-Load (ETL) process can be time-consuming, but is one of the most critical features of working with data.
- Visualize data using modern principles and effectively incorporate graphics into storytelling. Demonstrate mastery of basic data visualization techniques as well as being able to articulate the strengths and weaknesses of different graphical approaches. Students will be able to justify never using a pie chart again.
- Appreciate for the epistemological limits of typical data science approaches.
- Bring key technologies within data science to bear on real-world problems.