

# BMKT 670

## Applied Data Analytics

### Course Description

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

- **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 most of the 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 110 minutes.

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.

### Zoom Link

All classes will be delivered in person (in GBB 104) and via Zoom. Welcome to the future, remember to ask the valet about your flying car.

### 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 statistical insights in a business context.
4. Students will receive additional practice and instruction with data science technology including Python, R, and SQL.

### Required Materials

1. *Data Science from Scratch*, Joel Grus, O'Reilly.
2. *Weapons of Math Destruction*, Cathy O'Neil, Crown.
3. *Everybody Lies*, Seth Stephen-Davidowitz, Dey Street Books.
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.

## Assessment

We'll do a variety of assessment throughout the semester. You'll do a couple of projects working with data sets, a reflection paper on some reading, and regular programming assignments. Throughout the course students will be introduced to data sets and given self-contained assignments. These assignments will require basic programming skills in one of the languages reviewed in the course (typically Python or R).

Grades will be calculated based on the following table:

<b>A</b>	<b>93 or more points</b>
<b>A-</b>	<b>90-92.9 points</b>
<b>B+</b>	<b>87-89.9 points</b>
<b>B</b>	<b>83-86.9 points</b>
<b>B-</b>	<b>80-82.9 points</b>
<b>C+</b>	<b>77-79.9 points</b>
<b>C</b>	<b>73-76.9 points</b>
<b>C-</b>	<b>70-72.9 points</b>
<b>D</b>	<b>60-69.9 points</b>
<b>F</b>	<b>59.9 or fewer points</b>

### Class Participation: 10%

Students are expected to attend class either online or in-person and participate in class discussions. Missing more than five classes without an excuse will result in losing half the class participation credit.

### Wedge Assignment: 25%

You will do a data engineering project based on data from the Wedge Co-op, the largest co-op grocery store in North America. For the Wedge assignment, you will receive clear deliverables and traditional written feedback and a grade. **This assignment will be due TBA.**

### Building a Text-Based Data Set: 15%

You will build a data set based on text data acquired either through APIs or web scraping. The deliverable for this assignment will be a text-based data set (that can be used in the analysis for BMIS 625, Text Mining). You'll submit your code as well as a one-page description that will include the question you hope your data set will answer and descriptive statistics of your data set. **This assignment will be due by TBA.**

### Acquire & Analyze Assignment: 25%

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. **This assignment is due TBA.**

### Ethics Reflection Paper: 10%

This assignment will be a short (under 500 words) paper reflecting on ethical implications of one of our required books, *Weapons of Math Destruction*.

### Class Assignments: 15%

During the semester you will receive files with code that you must complete during and after class. Some of these you will turn in for a grade. You will receive a grade of one, two, or three. Three means

the code is outstanding, two indicates it is adequate, and one means that the code is not adequate. Grades for this portion of the class will be scaled so that a student receiving a two on every assignment will get 2/3 of the available credit.

## Class Format

Class format will primarily be hands-on work. Lectures will be delivered asynchronously via YouTube. 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.

## Moodle Organization

There's a ton of material in this course and, let's be frank, I'm not the most organized data scientist you've ever met. Which is saying something. To make it easier for you to follow along, I'm experimenting with a new organizational scheme on Moodle. Here are the key folders:

- **Zoom Link/Course Recordings:** Links to access the course on Monday and Wednesday evenings as well as the course recordings.
- **Upcoming Week:** This section holds material that you will need before the next Monday class. If there are lectures, watch them. If there's reading, read it. If there is code, get that code running on your machine and follow any in-line instructions<sup>1</sup>. Note that your code is also available at our [github repository](#).
- **Previous Week:** Material moves from "upcoming" to "previous". So, if you failed to watch, download, or work with anything from the last week, here's where you can find it.
- **Course Slides:** All the slides from the course, as well as a link to the YouTube Playlist of recorded lectures.
- **Supplementary Materials:** Reading, videos, and other materials for the class.
- **Wedge Assignment:** The Wedge assignment is a beast, so we'll store materials related to it here.
- **SQL:** Any SQL-related material that...doesn't fit in the other categories. I'm starting to seem repetitive.
- **Week N:** Each week, when I move content from "upcoming" to "previous", I'll create a section that duplicates the material numbered by weeks. I'm hoping that will make it easier to find materials.

## 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

---

<sup>1</sup> Note: sometimes I'll include code that has "solutions" in the name. You *don't* need to get that code running before class.

<b>Week 3</b>	Data Management	Google Big Query & SQL	GBQ, SQL
<b>Week 4</b>	Data Management	Python	Data Cleaning
<b>Week 5</b>	Data Acquisition	Python	APIs
<b>Week 6</b>	Data Acquisition	Python	Web Scraping
<b>Week 7</b>	Data Acquisition	Python	TBD
<b>Week 8</b>	Data Manipulation	R	dplyr, ggplot
<b>Week 9</b>	Data Analysis	R	Regression, Logistic Regression
<b>Week 10</b>	Data Analysis	R	ANOVA, Clustering
<b>Week 11</b>	Data Analysis	R	Shiny Apps
<b>Week 12</b>	Data Analysis	Python	Agent-based Models
<b>Week 13</b>	Data Analysis		
<b>Week 14</b>	Thanksgiving		
<b>Week 15</b>	Working Sessions		

You may have noticed that we meet for four hours a week for a three-credit class. That gives us some flexibility in class meeting.

## 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. Good talk.

## 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.

## 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 campus 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.