University of Montana

ScholarWorks at University of Montana

University of Montana Course Syllabi, 2021-2025

Fall 9-1-2022

BMKT 670.V60: Applied Data Analytics

John Chandler
University of Montana, Missoula, john.chandler@umontana.edu

Recommended Citation

Chandler, John, "BMKT 670.V60: Applied Data Analytics" (2022). *University of Montana Course Syllabi, 2021-2025*. 1301.

https://scholarworks.umt.edu/syllabi2021-2025/1301

BMKT 670 Fall 2022
Applied Data Analytics

John Chandler, PhD
e: john.chandler@business.umt.edu
p: 406-544-8720
CRN 72020

Class meeting: Monday 4-6 in GBB 205 or [online](). Lab Thursday 4-6 [online only, (same link)]() .
Office Hours: Monday 2-3 and Wednesday 10:30-11:30 at the same [link]().

Welcome!

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

- **Structured Query Language (SQL):** SQL is, simply, the way people interface with regular data. Being good at SQL is a pre-requisite for getting an analytics job and succeeding in a data-intensive job.
- **Data Management**: How do we store data for easy retrieval? How do you clean data to prepare it for analysis? Managing data is often 80% of a data project and this calls will help you get better at this.
- **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 cover some techniques that are not covered in other parts of your MSBA curriculum.

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. These sessions are optional. For those of you who cannot attend these Thursday sessions live, please 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.
- **Working Groups** This section describes some group work you'll be required to do at the beginning of the semester.
- **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. We will begin the semester in person in GBB 205 and online at https://umontana.zoom.us/my/chandler.
- **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. We will meet online at https://umontana.zoom.us/my/chandler.
- **Office Hours:** I strongly encourage you to take advantage of our scheduled office hours to get help with concepts and assignments. I'm happy to meet at other times if the scheduled times don't work for you. If you can't attend office hours via Zoom for some reason, please let me know.
- **Moodle:** All file sharing will take place via Moodle[1].
- **Teams:** Teams will be our communication tool for the semester. We're going to use it to collaborate and it's a great way to get help from me and from your peers. You're a member of the UMT Analytics team (let me know if not). Please join the Applied Data Analytics channel within that team.
- **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 Teams. Reach me on email via 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 become proficient at SQL.
2. Students will learn how to manage a data set and prepare it for analysis.
3. Students will learn how to use programming to acquire data sets.
4. Students will learn to embed analytical insights in a business context.
5. Students will receive additional practice and instruction with data science technology, especially Python.
6. Students will be conversant in the key themes in the ethics of data science.

---

[1] 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.

Materials

1. (Required) *Become a SELECT Star*, Julia Evans, https://wizardzines.com/zines/sql/.
2. (Required) *Principles of Data Wrangling*, Rattenbury, Tye, et al. O'Reilly Media, 2017. https://www.oreilly.com/library/view/principles-of-data/9781491938911/
3. (Required) *Weapons of Math Destruction*, Cathy O'Neil, Crown.
4. (Required for A grade) *Uncanny Valley*, Anna Weiner, MCD or Picador Paper.
5. (Optional) *Everybody Lies*, Seth Stephen-Davidowitz, Dey Street Books.
6. (Optional) *Data Science from Scratch*, Joel Grus, O'Reilly.
7. (Optional) *Algorithms for Data Science*, Steele, Chandler, and Reddy, Springer.
8. (Optional) *The Statistical Sleuth,* Ramsey and Schafer, Brooks/Cole Cengage Learning, 2013.

The first three books are required for everyone. The third book, *Uncanny Valley*, is required for students contracting for an A grade. *Everybody Lies* is sometimes required for this course is a readable exploration of the types of analyses that are unlocked via the marriage of disparate data sets. *Data Science from Scratch* is a great book with Python implementations of common data science tasks. *Algorithms for Data Science* is, in some ways, an extension of Joel's book. *The Statistical Sleuth* is my recommended reference book on introductory statistics and you will not need it for most work in this course, but it can be useful in your library.

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 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. By the beginning of week three of the course, you'll choose your workload, your deadlines, and your desired grade.

It's difficult in a new class to determine the appropriate workload. At Week 7 you'll be free to contract for a different grade if you'd like.

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.
- 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^2.

- Everything from "Passing".
- Reading *Weapons of Math Destruction* and writing the reflection paper.
- Completing the quizzes for the SQL module at a 90% pass rate.
- Completing the five assignments for the SQL module at a 70% pass rate.
- Completion of the "Wedge Data Engineering" assignment at the C level.
- Three class assignment 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.
- Completing the quizzes for the SQL module at a 90% pass rate.
- Completing the five assignments for the SQL module at an 80% pass rate.
- Completion of the Wedge Data Engineering project at the B level.
- Five class assignment submissions.

Grade: A

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.
- Reading *Uncanny Valley* and attending a book club on the book.
- Completing the quizzes for the SQL module at a 90% pass rate.
- Completing the five assignments for the SQL module at a 90% pass rate.
- Completion of the Wedge Data Engineering project at the A level.
- 7 class exercise submissions.

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

---

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

	Grade		
	A	**B**	**C**
Item			
Books			
Weapons of Math Destruction	1	1	1
Become a SELECT Star	1	1	1
Principles of Data Wrangling	1	1	1
Uncanny Valley	1	0	0
SQL Module			
Weekly Module Quizzes	90%	90%	90%
Weekly Assignments	90%	80%	70%
Wedge Data Engineering			
Project & Write-up	1	1	1
Class Assignments			
Weekly Assignments (Wks 5-15)	7	5	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 four books assigned for this semester, though only students going for an A will be required to read all four. 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, *Become a SELECT Star,* is a short zine that you will purchase in electronic form and this book supports the first five weeks of the course. Weeks 6-12 we'll work out of *Principles of Data Wrangling*, another book available in electronic form. Students contracted for an A will read *Uncanny Valley* and participate in a book club meeting about it. For all the books, you will select your due dates in your contract form.

SQL Modules

The first five weeks of the semester will be dedicated to learning and refining our SQL knowledge. There will be two ways your work will be evaluated during this time.

1. **Quizzes**: Each of the first five weeks of the course will have a quiz that covers the reading, lectures, and other materials from that week. You must get at least 90% on the quiz, but you are allowed unlimited re-takes (the questions will vary between attempts). You should pass the quiz by the end of each week.
2. **Assignments**: Each of the first four weeks of the course will have an assignment asking you to write SQL queries for different tasks. You will submit your attempt by the end of the week. Your submissions will be marked correct or incorrect. The grade necessary to pass depends on your contracted grade (90% for A, 80% for B, 70% for C). If your first submission is lower than this score then you must resubmit your assignment until you receive a passing score.

Class Assignment Submissions

Many weeks we'll do work in class that supports the material covered in lectures or readings. Starting in week 5 you will receive programming assignments, typically in Python. These assignments will have starter code and ask you to complete various tasks. You can choose the assignments you want to complete 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[3].

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.

A Note on Collaboration

In a scholastic setting that is also preparing us for real world collaboration, the boundary between sharing information and plagiarism can be murky. All of the following statements are generally true:
- Data science is fundamentally collaborative.
- I want you to work together.
- Struggling with code is how we learn.

Given all of this, here are the guidelines I propose for the semester:

> For your assignments, no copying code written by other students. You're welcome to collaborate on developing the code, but don't copy and paste from one person's work to another's. You must be able to explain every line of code copied from the internet or that someone else helped you write. For write-ups and interpretation, do those independently, though they can be based on conversations with your collaborators.

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 Teams channel.

---

[3] This is subject to change. I'm trying to bend GitHub Classroom to my will, with limited success so far.

I've created a channel in our team called Beautiful Mistakes. During the semester I invite you to post to that channel as much as you like. You might also enjoy reading some of the submissions from last fall, since many deal with problems you might encounter. A post should look something like this:

- What you were trying to do.
- Your original code.
- As clear an explanation as possible 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. All posts must be submitted two weeks before the final class and you can't post more than two in any one week.

Wedge Data Engineering Project

Your "big" individual assignment 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 middle portion of the semester (Weeks 5-11 or so) and is required for all contract grades, but with differing levels of work by grade. The brief for the project holds the descriptions of the levels.

Data Project #2

We will work on a data engineering project together in the final weeks of the semester, if all goes according to plan. There will not be a graded portion of this project.

Working Groups

It is very difficult to be successful in this class working alone. I *strongly* encourage you to form a group to work with during the semester. Find people whose schedules are copacetic with yours and get on a regular working routine. The course, and the degree, will be vastly more enjoyable.

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 and will always be online. 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.

Guest Visits

During the course, we'll have a number of visits from people connected to data engineering. This will give us an opportunity to learn from some talented people in the industry. Toward the end of the class we'll conduct some mock interviews with outside guests so that we can learn a bit more about applying for jobs in data science and data engineering.

Course Outline

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

	Theme	Languages	Topics
Week 1	SQL Module	SQL	Basics of SELECT statements
Week 2	SQL Module	SQL	Some more advanced SQL language features
Week 3	SQL Module	SQL	An introduction to some of the fancier functions in SQL and Google BigQuery
Week 4	SQL Module	SQL	Operations other than SELECT such as CREATE, UPDATE, INSERT, and DROP
Week 5	SQL Module	SQL & Python	Getting Python to talk to SQL data
Week 6	Data Management	Learning Python	Data Science Introduction, Python Overview
Week 7	Data Management	Python	APIs, Python & SQL
Week 8	Data Acquisition	Python	APIs Part 2
Week 9	Data Management	Python	Working with Text Files
Week 10	Data Acquisition	Python	Data Engineering tasks in Python
Week 11	Data Acquisition	Python	Wedge Work
Week 12	Data Acquisition	R	`dplyr,ggplot`
Week 13	Thanksgiving Working Session	NA	Online working session
Week 14	Data Analysis	R	K Nearest Neighbors, Clustering
Week 15	Data Analysis	R	Principal Components Analysis

Code of Conduct

Compliance with COVID protocols set by the University is mandatory.

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 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 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. 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, 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/triosss/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 or 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.

MS in Business Analytics – Learning Goals

1. Knowledge and Application:
   - An understanding of a range of analytical and programming techniques
   - Ability to apply appropriate techniques to solve a variety of business/organizational problems
2. Communication:
   - Ability to effectively communicate data analytics results and translate into business decisions.
   - Ability to effectively use data visualization techniques.
3. Ethics/Data Stewardship:
   - An understanding of ethical implications of data stewardship and privacy.
4. Innovation:
   - Ability to harness data analytics to identify new sources of value and to reveal innovative insights.

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

- Understand database structure and write queries to access that data.
- 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.