Statistics II: Data Analytics – Spring 2026

Course Information

Course Number: QM 2023
Course Title: Statistics II: Data Analytics
Semester: Spring 2026
Sections (two):

- **QM-2023-01:** M/W 2:00 PM – 2:50 PM, Helmerich Hall 100 (Friday - Homework/Assignment Days)
- **QM-2023-02:** M/W 10:00 AM – 10:50 AM, Helmerich Hall 100 (Friday - Homework/Assignment Days)

Term dates: 1/20/2026 – 5/16/2026
First class meeting: Wednesday, January 21, 2026
Last regular class meeting: Monday, May 4, 2026
Calendar note: Spring Break is 3/16–3/20 per the university calendar. We will not hold class meetings that week. Deadlines and any asynchronous work are posted in Blackboard.

Instructor Information

Instructor: Dr. Cayman Seagraves, Ph.D.
Office: Helm 122-D
Phone: (918) 631-2541
Email: cayman-seagraves@utulsa.edu
Office Hours: Monday & Wednesday, 3:00 PM – 5:00 PM, and by appointment
Website: www.caymanseagraves.com

Course Description

This course provides a comprehensive, practical introduction to modern statistical and data analytics methods with an emphasis on real-world applications in business, finance, economics, and real estate. Rather than treating statistics as a collection of isolated techniques, this course develops **regression as a unifying framework** for estimation, inference, and causal analysis.

The course is designed to bridge the gap between academic theory and professional practice. You will move beyond spreadsheet-based analysis toward reproducible, code-driven workflows using Python—the language of contemporary data science and financial analytics. Through hands-on labs, real datasets, and a semester-long capstone project, you will learn not just *how* to perform statistical analyses, but *why* specific methodological choices matter and *how* to verify and critique analytical results, including those generated by AI tools and agents.

Key themes:

- Interpretation over pure computation
- Transparency and reproducibility
- Working with real, imperfect data

- Responsible integration of AI-assisted analysis

Learning Objectives

Upon successful completion of this course, you will:

1. **Master foundational statistical concepts** through computational practice: hypothesis testing, confidence intervals, and the foundations of regression analysis.
2. **Build and interpret regression models** ranging from simple bivariate relationships to complex multivariate specifications with interactions and non-linear terms.
3. **Apply econometric techniques** including fixed effects models, difference-in-differences estimation, and time-series forecasting to address causal inference problems.
4. **Develop practical data engineering skills** using Python: loading, cleaning, merging, and reshaping datasets for analysis.
5. **Construct robust analyses** by diagnosing violations of statistical assumptions, implementing corrective measures, and understanding when and why standard errors must be adjusted.
6. **Communicate results professionally** through publication-ready tables, visualizations, and written summaries that translate technical output into actionable insights.
7. **Think critically about AI** in data analysis: use AI tools and agents productively while auditing their outputs and taking full responsibility for the work you submit.

Required Materials

Software (Free & Open-Source)

We will use **VS Code for everything**, with two supported paths. The best overall option is GitHub Codespaces (cloud-based), which includes all necessary Python environments. A local option is available by using VS Code on your own machine.

CRITICAL: DO NOT INSTALL ANACONDA. Anaconda is NOT supported in this course. We use a pure Python workflow managed through VS Code.

GitHub & Student Benefits (Mandatory) A **GitHub account** is required for all course activities. You must also apply for the **GitHub Student Developer Pack** to receive free access to GitHub Pro, additional Codespaces hours (180 hours/month), and other professional tools.

- **GitHub Account:** Sign up for GitHub
- **GitHub Education:** Apply for Student Benefits (requires proof of enrollment/student email)

Option A: GitHub Codespaces (Best Overall, Zero Setup)

- **GitHub account** (Join our GitHub Classroom to access assignments)
- **VS Code in the browser** via GitHub Codespaces. This environment is pre-configured with the correct Python version and all required libraries. No local installation is required.

Option B: Local VS Code + VS Code Python Extension

- **Visual Studio Code** download
- **VS Code Python Extension:** Install directly via the Extensions view in VS Code.
- **Python Version:** When you open a `.py` file, VS Code will prompt you to install a modern Python version (3.12+). You **do not** need a separate distribution like Anaconda.
- **GitHub Desktop or Git (recommended):** VS Code uses Git for sync and submissions. VS Code source control overview

All necessary Python libraries (pandas, statsmodels, linearmodels, pyfixest, matplotlib, seaborn, etc.) are pre-configured in our GitHub Codespace environment. For local setups, we will use VS Code's extension-led installation.

Textbooks & References

No required textbook. The course draws from research-grade datasets and modern software documentation. The following resources are optional and recommended, with priority on free materials and Python-friendly content:

Recommended Core (Free)

- **OpenIntro: Introduction to Modern Statistics (IMS)** — Free online text with web and PDF options.
  OpenIntro IMS book page
  OpenIntro IMS web book
- **Python for Data Analysis (McKinney)** — Open access web book; strong pandas workflow and modeling examples.
  Python for Data Analysis
- **Econometrics-With-Python (weijie-chen)** — Python notebooks covering OLS, diagnostics, panel data, and time series.
  Econometrics-With-Python repository
- **EconometricsSlides (Nick Huntington-Klein)** — Open lecture slides for undergraduate econometrics topics.
  EconometricsSlides site
- **The Mixtape (Cunningham)** — Free online causal inference text and supporting materials.
  The Mixtape site
  Mixtape Sessions repository

Additional Reference (Optional, Not Required)

- **Wooldridge, J. M. (2013).** *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning. *(Reference for regression theory and diagnostics)*
- **Angrist, J. D., & Pischke, J. S. (2009).** *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. *(Reference for causal inference and panel data)*

All assigned readings and tutorial materials will be provided free online.

Hardware

A laptop (Windows, Mac, or Linux) is **required**. Since we primarily use GitHub Codespaces, you do not need to run Python or Git locally. However, a laptop with a physical keyboard and a full web browser is necessary; tablet-only access will not be sufficient.

Grading Structure

Your final grade reflects mastery across data engineering, econometric reasoning, and professional communication. Grades are awarded based on **1,450 total points** distributed across components below.

Component	Points	Weight	Details
GitHub Classroom Projects	550	38%	15 weekly assignments or mini-projects building technical and analytical skills. Points are verified via GitHub Classroom autograding and manual AI Audit reviews. All students follow the **absolute default REIT track** to ensure consistent, streamlined grading and peer support.
Weekly Online Quizzes (Blackboard)	150	10%	15 quizzes total (one per week). Administered via Blackboard as part of the asynchronous Friday workday.
Midterm Exam I (Week 8)	225	16%	In-class exam covering Weeks 1–7. All exam questions are aligned to the REIT default dataset.
Midterm Exam II (Week 11)	225	16%	In-class exam covering Weeks 8–11. Focused on panel data and causal inference.
Capstone Milestones (4 total)	200	14%	Semester-long empirical project using the REIT track. Graded incrementally at 50 points each (**M1** (Week 5), **M2** (Week 9), **M3** (Week 13), **M4** (Week 15)).
Final Presentation (Week 15)	100	6%	Team-based "Investment Committee" presentation. Graded on clarity, evidence, and professionalism.
Total	**1450**	**100%**	

Detailed Breakdown

1. **GitHub Classroom Projects (550 points)** *Assignment points are automatically verified via GitHub Classroom unit tests (autograding) and manual verification of AI Audit appendices. Assignment 08 is a no-submission review.*
    - **Assignment 01:** Environment Setup (25 pts)
    - **Assignments 02, 04, 05, 06, 07, 09, 11, 13:** Technical Projects (50 pts each)
    - **Assignment 03:** EDA & Visualization (50 pts)
    - **Assignment 10:** Fixed Effects (50 pts)

- **Assignment 12:** Time Series (50 pts)
- **Assignment 14:** Hackathon (10 pts)
- **Assignment 15:** Repo Final Check (10 pts)

2. **Weekly Online Quizzes (150 points)**
    - **Quiz 01 – Quiz 15:** 10 points each × 15 weeks = **150 points**
3. **Midterm Exams (450 points)**
    - **Midterm I (Week 8):** 225 points
    - **Midterm II (Week 11):** 225 points
4. **Capstone Milestones (200 points)**
    - **Milestone 1:** Data Pipeline (50 pts)
    - **Milestone 2:** EDA Dashboard (50 pts)
    - **Milestone 3:** Econometric Models (50 pts)
    - **Milestone 4:** Final Investment Memo (50 pts)
5. **Final Presentation (100 points)** *Team-based simulation in Week 15.*

Grading Scale

Grade	Points	Percentage
A	1305–1450	90–100%
B	1160–1304	80–89%
C	1015–1159	70–79%
D	870–1014	60–69%
F	Below 870	Below 60%

Assessment & Feedback Workflow

- **Blackboard quizzes:** Auto-graded, low-stakes checks administered weekly (immediate feedback). Completed as part of asynchronous Friday workday.
- **GitHub Classroom autograding:** Unit tests and run checks verify code correctness and reproducibility.
- **Peer review cycles:** Selected weeks include structured peer feedback on code and memos.
- **Instructor feedback:** Focused on capstone milestones, interpretation, and methodological rigor.

Course Policies

Attendance & Participation

Regular attendance is essential. Class sessions combine lecture, discussion, and live coding. Fridays are designated as an asynchronous workday: students complete the weekly quiz and make progress on the weekly assignment and capstone deliverables. Deadlines are posted in Blackboard.

- **Late arrivals:** Habitual tardiness will be addressed.
- **Asynchronous Friday workday:** Students are expected to complete the weekly quiz and advance their assignment/capstone work independently. Consistent failure to complete Friday deliverables will impact your quiz and assignment grades.
- **Excused absences:** University-approved activities, documented illness, family emergency, and jury duty qualify as excused. You must provide documentation (gold sheet from CSAS) in advance or as soon as possible.

Assignment Submission & Late Work

- **Deadlines:** All assignments are due by 11:59 PM on the stated date. Code assignments are submitted via our GitHub Classroom (set up in Week 1). Quizzes and select written submissions run through Blackboard.
- **Late submission policy:** Assignments submitted after the deadline will be docked 10% per day late, up to a maximum of 3 days. After 3 days, no credit is given unless an excused absence applies.
- **Exam makeup policy:** Missed exams can only be rescheduled for documented, university-excused absences. Notify me at least one week in advance if possible.

Academic Integrity

Academic dishonesty will not be tolerated. This includes:

- Submitting work that is not your own
- Plagiarizing code or analysis without attribution
- Allowing someone to copy your code or assignments
- Misrepresenting the role of AI tools in your work (see AI Policy below)

Consequences: Minimum penalty is a zero on the assignment plus a one-letter grade reduction in the course. Repeated violations will result in an F in the course and referral to the Dean of Students.

AI Tools & Responsible Use Policy

You are **encouraged to use AI tools**—including ChatGPT, Claude, GitHub Copilot, and similar platforms—in this course, provided you follow three core principles:

1. Disclose

Every assignment submission must include an **AI Audit Appendix** (1–2 pages) documenting:

- Which AI tools you used (and how)
- Specific prompts or requests you made
- What the AI produced
- How you verified or modified the output

Example: "*I used ChatGPT to debug a pandas merge operation. Error message: 'KeyError on ticker.' ChatGPT suggested using left_on/right_on parameters. I verified the output against a manual test case before integrating it into my script.*"

No appendix = No credit for the assignment. It's not optional.

2. Verify

You are 100% responsible for the accuracy and appropriateness of all work you submit. "The AI told me so" is an invalid excuse for errors in logic, methodology, or interpretation.

- If AI generates a regression output, verify that the model specification is sensible (e.g., no obvious multicollinearity, appropriate fixed effects).
- If AI writes explanatory text, cross-check it against course concepts and textbook definitions.
- Run diagnostic tests yourself; don't blindly trust AI-generated summaries.

3. Critique

A significant portion of your grade—especially on advanced assignments—comes from your ability to identify and correct AI mistakes. Some assignments will intentionally include "traps": scenarios where standard LLMs generate plausible-sounding but incorrect advice.

Example: "*The AI suggested a Random Effects model. However, the Hausman test p-value is 0.02 (< 0.05), so Fixed Effects is appropriate. I corrected the specification and re-ran the analysis.*"

Academic Accommodations

Students with disabilities should contact the Center for Student Academic Support (CSAS) to register for services and obtain formal accommodation documentation. CSAS provides:

- Testing accommodations (extended time, separate room, etc.)
- Note-taking services
- Tutoring and academic counseling
- Other services to support your success

CSAS Contact: Center for Student Academic Support
Phone: (918) 631-8433

Provide accommodation documentation to me at least one week before exams or assignments requiring special arrangements.

Classroom Conduct

This is a professional learning environment. Please:

- Silence phones during class (exceptions: legitimate emergencies).
- Avoid excessive side conversations; if you need clarification, ask me.
- Refrain from web browsing unrelated to the course; laptops should be used for coding and note-taking, not for email or social media.
- Treat classmates with respect; this is a collaborative learning space.

Behavior that disrupts the class (extended conversations, excessive phone use, hostile commentary) may result in a 5% deduction from your next exam grade.

Title IX & Sexual Misconduct

The University of Tulsa is committed to maintaining a safe, respectful community. Sexual misconduct—including harassment, assault, dating violence, domestic violence, stalking, and exploitation—is prohibited.

If you or someone you know has experienced sexual misconduct:

- **Title IX Coordinator:** (918) 631-4602
- **Office of Violence Prevention:** (918) 631-2324
- **TU Counseling & Psychological Services:** (918) 631-2241
- **Campus Security:** (918) 631-5555 (non-emergency) or 911 (emergency)
- **Domestic Violence Intervention Services:** (918) 585-3163 or (918) 743-5763

You have the right to report and access resources. Every student on campus deserves safety and support.

Syllabus Changes

This syllabus is a living document. While every effort is made to adhere to the schedule and policies outlined, changes may be necessary. Any updates will be announced in class and posted on the course GitHub repository. You are responsible for staying informed of announcements regardless of attendance.

Course Outline (Concise)

This syllabus provides the high-level roadmap. The **detailed, per-meeting schedule** lives in:

- `Week-by-Week-Schedule` (topics, M/W/F structure, labs)
- `QM2023-Spring2026-Week-Date-Map` (calendar dates, due dates, and university deadlines)

At-a-Glance Modules

Module	Weeks	Focus	Major deliverables
Module 1: Data Science Foundations	1–3	Python workflow, reproducibility, data wrangling, EDA/visualization	Weekly assignments
Module 2: Regression & Inference	4–7	OLS, multiple regression, diagnostics/robust SE, logistic regression	Weekly assignments; **Capstone Milestone 1 (Week 5)**
Module 3: Panel Data & Causal Inference	8–11	Panel structure, fixed effects, difference-in-differences	Weekly assignments; **Capstone Milestone 2 (Week 9)**
Module 4: Time Series, ML, & Capstone	12–15	Forecasting, prediction vs. inference, professional communication	Weekly assignments; **Capstone Milestones 3 (Week 12) & 4 (Week 14)**; presentations

Key Assessment Weeks (Week numbers)

- **Midterm Exam I:** Week 8 (covers Weeks 1–7)
- **Midterm Exam II:** Week 11 (covers Weeks 8–11)

- **Capstone milestones:** Week 5 (M1), Week 9 (M2), Week 12 (M3), Week 14 (M4)
- **Final presentations:** Weeks 14–15 (Fri, May 1 & Mon, May 4; team-based "Investment Committee" format)
- **Final exam:** No cumulative final exam (see week-by-week schedule for finals-week expectations)

Note: Spring Break is 3/16–3/20 (no class meetings). Specific assignment due dates and any asynchronous work are posted in Blackboard and mirrored in the week↔date map.

Semester-Long Capstone Project

Overview

Rather than disconnected problem sets, you will work on a **semester-long capstone project** that simulates real-world quantitative analysis.

Format (Hybrid): You will work in a small team for the shared analysis workflow (shared repo, shared figures/tables), while still completing **individual accountability** deliverables (reflection and peer evaluation). The due dates and milestones remain the same.

Research Prompt

"*You have been hired as a Junior Quantitative Analyst at a real estate investment firm. The Investment Committee is concerned about REIT return sensitivity across sectors in the aftermath of the 2022–2023 Federal Reserve interest rate hikes. You must build a data pipeline to analyze historical sensitivity of REIT returns to interest rates and estimate the causal impact of the rate hike cycle on different sectors.*"

Data & Sourcing

The course provides a high-quality, research-grade dataset as the default for all work.

Default Track: REIT Master + Factor Panel

1. **REIT Data:** REIT Master Panel – monthly returns and characteristics for US REITs. Provided by instructor.
2. **Economic Data:** FRED (Federal Reserve Economic Data):
    - FEDFUNDS (Federal Funds Rate)
    - CPIAUCSL (Consumer Price Index)
    - DFF (Effective Federal Funds Rate)
    - Other indicators of your choice
3. **Panel Structure:** Rows = REIT permnos; Columns = Monthly time periods (60+ months of data)

Alternative Datasets (Capstone Only by Exception): Students may request to use Orbis academic data or an open dataset from the **Open Data catalog** (`OpenData_rows`) for their capstone project only. This requires a formal proposal and instructor approval by Week 4. Proposals are granted by exception only; if not approved, students must use the default REIT track. See `Dataset-Options-Guide` for requirements.

Milestone Breakdown

Milestone 1: Data Pipeline (Due Week 5, 50 points)

- **Deliverable:** Python script (.py or .ipynb) that programmatically fetches, cleans, and merges datasets

- **Requirements:**
  - Load REIT Master data (500+ permnos, 60+ months)
  - Fetch FRED economic data (MORTGAGE30US, UNRATE, CPIAUCSL, etc.)
  - Reshape to long format panel structure (Entity=REIT/Permno, Time=Month)
  - Handle missing values with defensible choices (document in comments)
  - Merge datasets on Date/Month, maintain data integrity
  - Save output as tidy CSV with metadata
- **Rubric:** Reproducibility (15 pts), code quality (12 pts), data structure (15 pts), documentation (8 pts)

Milestone 2: EDA Dashboard (Due Week 9, 50 points)

- **Deliverable:** Jupyter notebook with exploratory analysis and visualizations
- **Requirements:**
  - Correlation heatmap: interest rates vs. REIT returns across permnos
  - Lagged effect analysis: how long does a rate change take to affect returns?
  - Sector segmentation: identify "sensitive" (volatile) vs. "resilient" (stable) REIT sectors
  - Time series decomposition: extract trend, seasonality, residuals
  - Summary statistics by permno and time period
- **Rubric:** Data fetching (10 pts), visualization quality (20 pts), analysis depth (15 pts), presentation (5 pts)

Milestone 3: Econometric Models (Due Week 12, 50 points)

- **Deliverable:** Regression results, diagnostic tests, robustness checks, interpretation memo
- **Requirements:**
  1. **Model A:** Fixed Effects regression (controlling for REIT-level characteristics)
     - Specification: `Return ~ Factors + Economic_Indicators + REIT_FE + Time_FE`
     - Interpret: within-REIT sensitivity of returns to factor exposures
  2. **Model B:** Difference-in-Differences or Alternative Specification (e.g., sector shocks)
     - Specification: interaction of treatment and post-period
     - Causal effect estimation
  3. Robustness checks:
     - Clustered standard errors (at REIT/Permno level)
     - Alternative lag structures
     - Sensitivity to outlier treatment
     - Placebo tests (if using DiD)
  4. Publication-ready results tables with interpretations
- **Rubric:** Model specification (15 pts), diagnostics (12 pts), interpretation (18 pts), presentation (5 pts)

Milestone 4: Final Investment Memo (Due Week 15, 50 points)

- **Format:** Team memo + individual addendum (hybrid capstone)

- **Team Memo Deliverable:** Professional 5–7 page PDF report (team submission)

  - **Executive Summary (0.5 page):** 2-3 sentence finding + investment implication
  - **Methodology (1 page):** Data sources, panel structure, model equations, variable definitions
  - **Results (1.5 pages):**
    - Table 1: Fixed Effects model with factor sensitivity

        - Table 2: Robustness or DiD specification
        - Figure 1: Key visualization (e.g., factor vs. return trends)
        - Interpretation prose explaining coefficients in economic terms
    - **Conclusions & Recommendations (1 page):** Buy/Hold/Sell for specific markets
    - **References & AI Audit Appendix (0.5-1 page):** Data sources cited; AI Audit required per course policy

- **Individual Addendum (1 page):** Each student submits a short addendum that includes:
    - What you personally contributed (2–4 bullets)
    - One methodological decision you would defend (2–4 sentences)
    - One limitation you think matters most (2–4 sentences)
    - AI audit notes (if AI was used in your portion)

Grading (50 points total): Team memo (40) + individual addendum (10)

Evaluation Criteria (AI-Aware Rubric)

Dimension	Excellent (A)	Satisfactory (C)	Unacceptable (F)
Reproducibility	Code runs cleanly; dependencies listed; relative paths; all outputs regenerable	Code runs but requires minor debugging	Code fails to execute
Technical Rigor	Models economically sensible; assumptions checked; robustness explored	Model is reasonable but diagnostics are incomplete	Logic is fundamentally flawed
Interpretation	Explains *why* models were chosen; AI Audit identifies and corrects errors	Comments describe *what* the code does; interpretation is basic	No interpretation; AI outputs accepted blindly
Communication	Tables/figures publication-ready; narrative flows logically	Visuals functional but lack polish	Visuals misleading; narrative unclear

Course Resources & Support

Office Hours & Tutoring

- **Instructor Office Hours:** Monday & Wednesday, 3:00–5:00 PM in Helm 122-D, or by appointment
- **Course GitHub & Q&A:** Discussion forum for technical questions (first 24 hours response time)
- **Center for Student Academic Support (CSAS):** Free tutoring, academic coaching, study skills workshops
    - Center for Student Academic Support
    - Phone: (918) 631-8433

Technical Help

- **Python Help:** Official documentation links will be provided; we recommend Real Python and StackOverflow

- **Git/GitHub:** GitHub Desktop or command-line tutorials
- **Error Debugging:** Bring specific error messages to office hours; we'll work through them together

Library Resources

The University of Tulsa Library offers:

- Database access (Bloomberg Terminal, academic journals)
- Research consultations
- Statistical software tutorials

Visit the McFarlin Library website or ask a librarian.

Program Learning Objectives

This course contributes to the following University of Tulsa B.S.B.A. Program Learning Objectives:

1. **Apply financial analysis methods** to assess risk and return characteristics of investments (through regression, time-series forecasting, and capstone project)
2. **Communicate effectively** financial and analytical concepts using quantitative evidence (through EDA dashboards, technical writeups, and oral presentations)
3. **Demonstrate analytical reasoning** using data and modeling to support business decisions (through all assignments and the capstone)

Technology & Accessibility

Required Technology

- Laptop (Windows, Mac, or Linux) with a modern web browser
- Software: GitHub Codespaces (cloud environment); local Python/Git optional
- Reliable internet for accessing course materials and submitting work

Accessibility

If you require accommodations due to a disability, CSAS provides confidential support. Reach out early in the semester so accommodations can be arranged in time for your needs.

Final Notes

This course is designed to equip you with practical, market-relevant skills in data analysis and statistical modeling. The emphasis on Python, reproducibility, and critical thinking reflects how modern organizations actually conduct analysis—not how it was taught 20 years ago.

AI tools are here to stay. Rather than fighting that reality, we embrace it: use AI productively, but remain the human in the loop. You'll graduate not just with competence in a specific programming language, but with the judgment to know when to trust a tool—and when to doubt it.

Welcome to the course. Let's build something real.

Contact & Questions

If anything in this syllabus is unclear, please ask during office hours or via email. I'm here to support your learning.

Best way to reach me: Email cayman-seagraves@utulsa.edu (24-hour response target)

This syllabus is subject to change at the instructor's discretion.