

Final Project Regression Fall 2025

Due: Thursday 10/10 @ 11:59pm

The final project is designed to showcase and expand students' linear/logistic regression knowledge and test their ability to teach and communicate technical material to a broader audience.

Your task is to **1. write and publish a blog post** and **2. create and publish an accompanying interactive graphic or app** whose purpose is to educate a reader (e.g. a future MSDS student) and help them develop intuition about a topic in regression. You should focus on doing a “deep dive” into the topic – in other words, the scope your blog post/app must extend past the basic material we have covered in class and include outside references. ***Your group must discuss your topic with me to get approval before you get started to ensure your choice is well-suited for this project.***

You have freedom with where you want to go with your lesson— try to be creative and illustrate **your point of view**, not just what a textbook would say. In other words, ***show me something cool*** about your topic! Tell me **your thoughts** on the topic, how it can be used, what makes it interesting, what are its drawbacks, and so on. ***I am deliberately giving you room to explore the topic in an open and potentially unconventional way—if you just give me a boring, standard textbook lesson of the foundations without incorporating any personal perspective or depth, then you are missing the point!***

Submission:

Links to your published blog post and app, and either a notebook or Github with all your Python code & scripts. **NOTE: Your blog post and app must be live and accessible without local installation. This is part of building up your online data science portfolio! Warning: Streamlit apps tend to break/go to sleep, so they're not ideal.** Each group will have one submission. You will also submit a breakdown of work and peer review score for each of your project partners individually and confidentially.

Blog post & app description:

Your final project submission should have the following components, at minimum:

- A clear and motivating introduction to the topic which is accessible to a wide audience (in other words, don't jump straight into technical details).
- A discussion and illustration of the basics (i.e. what we've talked about in class) using either an example dataset or a simulation.
- At least one section extending the basics to more advanced or related aspects of the topic and illustrating them either through other data examples or simulations. You will need to do some independent research for this part!
- An interactive graphic/app which a reader can use to build their intuition and explore the topic, either with your data, their own data, or through simulation.
- A proper reference section to hold your sources.

Grading rubric:

Component	Points
Topic Intro & Motivation	/5
Lesson and Illustration of the Basics	/10
Topic Extensions (i.e. content beyond what we covered in class)	/25
Application to Dataset (Both basics + extension)	/20
Quality of Interactive Visuals/App	/25
References	/5
Collaboration Score (Peer Review: were you a good partner to work with?)	/10

****If blog/app is not published online and needs local installation: -10pt penalty**

Example Suggestion Topic Ideas (non-exhaustive):

- Generalized Linear Models
- Mixed Effect Models
- Nonparametric Regression
- Semiparametric Regression
- Generalized Additive Models
- Regression Splines and Nonlinear Regression
- Bayesian Regression Models
- Functional Regression Models (e.g. scalar-on-function, function-on-function)
- Object/Fréchet Regression (e.g. regression of covariance matrices, networks, shapes)
- Quantile Regression
- Parallel Computing & Time Optimization for Large-Scale Regression
- High Dimensional Regression
- Causal Inference in Regression
- Principal Components Regression + Partial Least Squares
- Robust Regression
- Spatio-temporal Regression
- etc.