DS 325 Final Project Report
Spring 2025, Prof Roth

Include this document as a **PDF** in the github repo with your project.

The final project report should answer the Why? What? and How?

- Why did you pursue this project?
- What did you accomplish? What is the take-home result/decision/tool etc?
- How did you go about this project? Which tools did you use and why?

Guidelines:

- 1000 words max (not including references, citations, figure captions).
- 2 figures minimum, at least one should show the results of a model (e.g. confusion matrix, predicted vs true values in regression, data clusters, etc)
- Cite any sources used including AI (and how AI was used). Many have selected data sets that have been released as alongside academic publications; be sure to cite these works.
- The paper should be well-written with correct spelling and grammar. I highly recommend getting the report proofread either by a fellow student, an instructor at the writing center, or using an AI writing assistant (e.g. Grammarly). The initial writing should be your own and you are responsible for any work submitted; if an AI agent plagiarizes, and you submit the work uncited, you've effectively plagiarized.

Format and approximate length suggestions:

Introduction/Abstract (200 words, 6-8 sentences)
This section summarizes the entire project. So, in addition to introducing and motivating your project, you will very briefly describe your results and conclusions in this section.

- 1 or 2 sentences on the background of the topic of your project.
    - Example: "The NFL draft is a three day event teams sequentially select college players to join their organizations. Generally, players are selected based on numerous factors (statistics at the collegiate level, their physical attributes, combine performance, etc) with more athletic and more accomplished players selected earlier."
- 1 sentence to motivate your thesis
    - Example: "However, sometimes players with similar attributes are drafted at dramatically different positions in the draft order."
- 1 sentence thesis
    - Example: "I suspect that where a prospect played at the collegiate factors into teams' evaluations and subsequent placement in the draft."
- 2-3 sentences about your approach
    - Example: "Using data from the past 10 drafts and dividing players by position and conference (Power Four or other), I apply linear regressions to each position to predict draft selection order based on player statistics."

- 1 sentences about your results
    - Example: “Inspecting the coefficients of regression, I find that for some positions—namely wide receiver and defensive end—the bias term is lower for Power Four than others, suggesting a preference for players from those conferences.”

    Note: The above example is fictitious. I did not perform this analysis.

Methods (300 words)
Here you’ll summarize your process.

- Describe the dataset(s) you used. Which features are you using? What is the prediction/goal?
- Did you do any significant data cleaning/preprocessing?
- Are you making assumptions that should be explained in advance of modeling?
- What model(s) did you use? Why this choice of model?
- Did you try or compare multiple models?
- This section does not have to be 100% prose. Consider how you would want to organize the information for readability:
    - A list of steps
    - Subsections for each stage
    - A table

What NOT to do:

- You should always refer to models and processes by their general names, not their function names. If you want to reference the function, do it in a parenthetical.
    - For example “We compared random forests (RandomForestClassifier) and gradient boosted trees (GradientBoostingClassifier), exploring different hyper-parameters using a grid search with cross validation (GridSearchCV).”
- You may mention the programming language and main packages you used, but you don’t need to explain to me what they are.
- Avoid “then we did…” “then we did…” “then we did…” repetition.

Results (200 words)
Results are the outputs and assessment of your model, without your interpretation.

- You should have at least one figure in this section showing how your model performed.
- Any assessments (confusion matrix, R^2, accuracy/recall/precision) belong in this section.
- If you are reporting multiple metrics, consider a table or list.

Discussion (300 words)
Here you’ll include any interpretations of your results and reflections on your process.

- How would you interpret your results? Do your results inform some decision or opinion? Can they be used in some way?

- Did you encounter any hurdles? How did you overcome them?
- What was successful? Is there something else you expected? Something else you should've/could've done?
- Does this project inspire future work? How would you continue?

Citations and Attributions (no limit)

- Cite the data set, both the URL and the individual/organization that hosts the data.
- Cite any publications that go along with the data or that you've referenced.
- Cite any use of AI.

Figures and Captions (no limit)

- Include at least 2 figures. Figures should have explanatory captions and should be referenced in the text.
- At least one figure should be in the Results section.
- I recommend placing Figures and Captions in tables (you can make table borders invisible). This will help with placement and alignment.

Code (not part of the report)

- Code must run.
- Once you have complete and functional code, edit out any extraneous portions (code where you were just figuring things out, code with errors, etc).
- Do not import packages in code more than once. If you are installing a specialty package in your code (e.g. !pip install ____ or !conda install ____), please comment out that line before submission.
- Code should be documented with markdown cells and comments.

Rubric

Section	Component	Comments (Present?/Notes on what to improve)
Front material		
	Title	
	Author Name(s)	
Introduction		
	Background	
	Motivation	
	Objective or Thesis	
	Summary	
Methods		
	Data description	
	Pre-processing	
	Modeling/Model Selection	
	Clear, concise, complete	
	Appropriate choice of approach	
Results		
	Appropriate and clear figures	
	Clear, concise, complete	
Discussion		
	Major conclusions, take-aways	
	Hurdles and solutions	
	Caveats or limitations	
	Future work	
Citations		
	Proper attribution to data and other resources	
	Citations for any AI used, including prompts.	
Figures		
	Figures/tables well formatted	
	Axes are labeled	
	Captions are short and useful	
Overall		
	Grammar and spelling	
	Clarity and organization	