CSC 485-001: Data Science Capstone
Spring 2026

The critical checkpoints for a project are as follows:

- **Statement of Question:** Students will develop 2-3 possible questions to explore throughout the semester. A brief explanation of each question will be given to give context to non-experts, and students will identify possible types of data that might be useful.
- **Identification of Datasets of Interest:** Students will choose one question of focus, and identify 2-3 datasets that will help them answer this question. Students will identify the steps necessary to acquire these datasets.
  - A good source of public data sets can be found here: `https://guides.library.msstate.edu/c.php?g=1023952&p=7417324`
  - **Note:** Projects using datasets with already published analysis (such as those found on Kaggle) will be expected to include analysis beyond what is already available through those sources.
- **Draft Data Audit:** Students will gain an understanding of the datasets they are using, with the goal of making sure that their data is reusable, and their work is reproducible. This is also a mechanism for understanding the origins of the data to help uncover bias. Precise text of the data audit assignment will be provided.
- **Data Draft Analysis:** Students will use the knowledge gained in data analysis from previous courses to create models and compelling visualizations of their data. Model results and visualizations will be accompanied by 100 word text.
- **Presentation:** Students will create a short 5-10 minute presentation highlighting key takeaways from the project and at least one illustrative data visualization. Students will present this slide in conversation with the class.
- **Project Paper:** Students will prepare a paper that specifically addresses their policy question. This paper will include a "Data Overview" section which contains a narrative version of the data audit a "Methodology" section that includes details of models used, a "Results" section that discusses model output and validation, and a "Conclusions" section that frames results in terms of the posed question.

Project Rubric:

Core competency	What Success Looks Like	Total Points
Data relevance and integration	Effectively integrates two or more data sources that are highly relevant to the question at hand.	4
Exploration of the data and its provenance	The source of the data is thoroughly discussed, and unresolved questions about the data collection and/or source are clearly identified.	2
Visual communication	The report uses a variety of different kinds of visualizations that follow principles of effective design and that display and highlight important and relevant features of the data that address the question. Moreover, the choice of visualizations are suitable for the aspects of the data being highlighted.	5
Data modeling	The report includes appropriately applied models for regression, classification, and/or clustering. The method and results of the modeling are clearly described, as are the limitations of the analysis.	5
Written communication	The report is clearly written, with written analysis that synthesizes and explains the visualization and modeling aspects of the report, and at an appropriate level of technicality.	5
Clarity and appropriateness of conclusions	The report uses the data to give a great deal of insight related to the chosen question. Uncertainty and limitations of the data and analyses in the report are clearly communicated. The report includes enough information about the context of the question so that a non-expert can clearly understand the issue and its importance.	2
Notebook(s) with code	Along with the report, student submits well-commented notebook(s) with code used for the project.	2
Presentation	Student gives a clear and engaging presentation that highlights the main results of the project, as well as any major challenges or limitations. Presentation includes descriptive title, one appropriate and compelling data visualization, and several key takeaways.	5