# Final Project Instructions for Behavioral Data Science Course

## Objective

The objective of this final project is to consolidate and apply the knowledge and skills acquired in the Behavioral Data Science course. Students are offered two options for completion: Data-driven and Model-driven.

## Across both options:

- **Models**: Students are required to apply three different models from the course. Standard linear and logistic regression will not count towards this requirement, but generalizability methods and regularization techniques will. Models may include but are not limited to ridge regression, artificial neural networks, trees, and cluster analyses.
- **Reporting**: All students must provide a comprehensive report detailing their methodology, data, analyses, results, and interpretations.
- **Coding Standards**: Code must be well-annotated and adhere to best practices for clarity and reproducibility.

## Option 1: Data-driven Project

In this option, students will source their own dataset or choose from publicly available data. Students must ensure that the data chosen aligns with the themes and models relevant for behavioral data science.

### Requirements:

1. **Data Description**: Provide a detailed description of the dataset including the origin, variables, sample size, and any cleaning or preprocessing steps taken.
2. **Model Application**: Apply three models from the course on the dataset. Include model diagnostics, validation, and any measures used to ensure the robustness of the model.
3. **Interpretation**: Offer a thorough interpretation of the modeling results, including implications for behavioral science.
4. **Ethics**: Discuss any ethical considerations related to data handling and analysis.
5. **Documentation**: Code should be submitted along with the report, documented to explain the process flow clearly.

## Option 2: Model-driven Project

For this option, students will engage in one or more Kaggle competitions that are in line with behavioral data science. This is to foster an understanding of competitive model building and to improve practical skills in machine learning.

**Requirements:**

1. **Kaggle Competition Selection**: Select a Kaggle competition(s) that involves behavioral data or has relevance to behavioral analyses.
2. **Submissions**: Make at least 10 different submissions to the competition(s), documenting your attempts, changes, and progress.
3. **Evidence**: Provide evidence of your submissions, which may include links to Kaggle Notebooks or the R scripts and their outputs.
4. **Reflection**: Reflect on the learning process, including what strategies were attempted, what worked, what didn't, and why.
5. **Code and Report**: Submit all code along with a detailed report that includes insights gained from the competitive process.

## Final Report - Common Components

Both options require a final report with the following components:

- **Introduction**: Brief overview of the chosen data or competition, and objectives of the analysis.
- **Methodology**: Explain your workflow from data preparation to model selection and validation.
- **Results**: Present the outcomes of your analysis/modeling in an organized manner, using visualizations where appropriate.
- **Discussion**: Provide insights and interpretations of the results in the context of behavioral data science. Address potential limitations and suggestions for future work.
- **Conclusion**: Summarize the key findings and learnings from the project.
- **References**: Cite all sources, including packages, articles, and any resources used.

## Submission

Submit the report alongside code and any supporting documents as a compressed file (.zip) through the designated submission platform before the final deadline as specified in the course schedule. Ensure that reports are written with a clear structure, well-organized sections, and are proofread for clarity and correctness.

---

## Grading

Grades will be based on the complexity and appropriateness of the chosen models, the depth of analysis and interpretation, the quality of the final report, code standard, and adherence to ethical standards in data handling and analysis. Participation in Kaggle for the model-driven option should demonstrate iterative improvement and learning.

## Important Final Notes:

All projects are expected to adhere to the principles of open science and reproducible research. Ethical considerations and the Vanderbilt Honor Code principles apply. The use of Artificial Intelligence to assist in your work is permitted and encouraged; however, the integrity and originality of your final submission must be maintained.

---

## Deadline: 04-30-2024 at 11:59:59pm

Submit all materials via Brightspace before 11:59:59pm on the due date. Late submissions will be subject to the late work policy outlined in the syllabus (80% for up to one week late).

Good luck, and let your analytical creativity thrive!