**DT 320: Into to Data Science Final Project**

*Spring 2024*

**Due Date:** Friday, May 3rd by 5:00 PM.

**Overview**:

Projects are integral components of this course, providing opportunities to apply data science concepts to real-world scenarios. To ensure a successful and collaborative project experience, the following policies are in place:

1. **Team Formation:** Projects may be conducted individually or in teams, as specified by the instructor. Teams should be formed early, and collaborative dynamics should be maintained throughout the project duration.
2. **Project Proposal:** A project proposal outlining the objectives, scope, methodology, and expected outcomes must be submitted and approved by the instructor before the project commences. This ensures alignment with course objectives and provides guidance for successful project execution.
3. **Progress Updates:** Regular progress updates, including milestones achieved and challenges faced, are required. These updates foster transparency, allow for timely feedback, and ensure that projects are on track.
4. **Collaboration and Communication:** Effective communication within teams is crucial. Utilize designated channels for team communication, and promptly address any issues or concerns that may arise during the project.
5. **Documentation:** Thorough documentation of the project, including data sources, data cleaning, manipulation, mining, visualization, interpretation, and code, is essential. Clear documentation enables understanding, transparency, and reproducibility.
6. **Final Presentation:** Each team is required to deliver a final presentation on May 03, 2024 from 10:15 AM to 12:15 PM, showcasing the project's objectives, methodology, findings, and visualizations. This presentation provides an opportunity to communicate results effectively.
7. **Individual Contribution:** In team projects, individual contributions should be clearly delineated. Each team member is expected to actively participate and contribute to the project's success.
8. **Ethical Considerations:** Projects must adhere to ethical standards, respecting privacy, confidentiality, and legal requirements. Any ethical concerns should be addressed promptly.
9. **Submission Deadline:** Projects must be submitted by Friday, May 3rd by 5:00 PM. Late submissions may be subject to grade deductions.

**Dataset Sources:**

The following links provide numerous datasets (though you are free to use your own):

* **Kaggle**: kaggle.com
* **UCI Machine Learning Repository**: <https://archive.ics.uci.edu/ml/index.php>
* **University of Florida Statistics**: <http://users.stat.ufl.edu/winner/datasets.html>

**How to document your project:**

Documenting a data science project effectively is crucial for ensuring clarity, reproducibility, and transparency. Here's a guide on how students should document their projects:

* + - 1. **Project Overview:**
* Begin with a brief summary of the project objectives, scope, and significance.
* Clearly state the problem statement or research question the project aims to address.
  + - 1. **Data Sources:**
* Provide details about the datasets used, including their origins, formats, and any preprocessing steps applied.
* Include links to the datasets or references to where they can be accessed.
  + - 1. **Data Cleaning and Preprocessing:**
* Document the steps taken to clean and preprocess the data, such as handling missing values, outliers, and data transformations.
* Include code snippets or scripts used for data cleaning and preprocessing.
  + - 1. **Data Analysis and Modeling:**
* Describe the methodologies and algorithms used for data analysis and modeling.
* Explain the rationale behind selecting specific algorithms and parameters.
* Include code or scripts for model training, validation, and evaluation.
  + - 1. **Visualization and Interpretation:**
* Present visualizations of the data, model outputs, and analysis results.
* Provide interpretations of the visualizations and discuss insights gained from the analysis.
  + - 1. **Results and Evaluation:**
* Summarize the key findings and results of the project.
* Evaluate the performance of the models and discuss any limitations or areas for improvement.
  + - 1. **Code Documentation:**
* Ensure that code is well-documented with clear comments explaining its purpose, inputs, outputs, and functionality.
* Use consistent coding conventions and formatting for readability.
  + - 1. **Reproducibility:**
* Include detailed instructions on how to reproduce the results, including software dependencies, libraries used, and configuration settings.
* Provide scripts or notebooks that automate the data preprocessing, analysis, and modeling steps.
  + - 1. **Final Presentation Materials:**
* Include materials used for the final presentation, such as slides or visual aids.
* Summarize key points and insights from the presentation for reference.
  + - 1. **Acknowledgments and References:**
* Acknowledge any individuals, organizations, or resources that contributed to the project.
* Provide references to relevant literature, datasets, and external sources used in the project.