Review: Final Team Project Instructions

The final project will assess the application of data mining/data science methods applied to solve real-world business use cases introduced in this course. You will be assigned to a group by your instructor. You and your teammate(s) will work closely to find a business dataset, clean and pre-process it, and perform a data science project on the data using Python. Final projects and datasets from prior courses cannot be used. It is **REQUIRED** that you and your team use GitHub as a code hosting platform to manage version control and collaboration during this project. It is also necessary to create and add a README file to your GitHub repository. Building web applications (using Streamlit or Python's Dash or similar) is a *big advantage* and *valuable practice* but not a requirement. It is recommended that you follow the PEP 8 – Style Guide for your Python code in the final project.

Project Timeline:

- Module 2 (by the end of Week 2): The course instructor will group students into teams of two to three members.
- Module 3 (by the end of Week 3): Each team should select and introduce a dataset. The team representative will need to submit the "Team Project Status Update Form."
- Module 6 (by the end of Week 6): Each team should submit the following deliverables for the course project in the final week:

1.
    1. **Video** - One 15-20 minute video presentation by all team members. Your video should address two audiences, thus be divided into 2 parts. Submit one mp4 file.
        - Part 1: Technical (peer data scientists) - Video should cover your technical notebook.
        - Part 2: Non-Technical (Business Executive) - Using PowerPoint or other tools to tell your data story that includes recommendations and/or next steps.
    2. **Business Brief** - One PDF document containing your business brief.
    3. **Technical Notebook** - One PDF document of your technical notebook containing your project code with thorough code notation and explanations of approach, code output, interpretation of your code output, and actionable insight recommendations and recommended next steps.
    4. **GitHub Repository** - Containing your project technical notebook(s) that demonstrates team collaboration. In your final technical notebook submitted on Blackboard, include a link to your GitHub repository.
    5. **Presentation (PowerPoint/Others)** - Present your "Data Story." One PDF or PPT document (at most 6 slides) containing slides prepared for a non-technical executive audience, with an executive summary of the project, your actionable insights, your business recommendations, call to action, and next steps for future work.

**It is critical to note that no extensions will be given for any of the final projects due dates for any reason, and final projects submitted after the final due date will not be graded.

Project Datasets:

You will be responsible for finding a *business-related* dataset for this final team project. Several free data resource links are provided within the Resources page on the MS-ADS GitHub website.

Requirements:

Divide the work equally between the team members for the following steps, and everyone needs to code. You are expected to write high-quality, efficient, and readable code in Python. This project requires that you and your team create a business brief, conduct a video presentation, submit a technical notebook with a link to your GitHub repository, and an executive summary presentation using PowerPoint.

For your video presentation:

- Give a 15-20 minute presentation, the first part covering your technical report and the second part covering your PowerPoint presentation.

- Include a technical explanation of:
    - The clearly defined problem statement and data description
    - Exploratory Data Analysis (EDA)
    - Data pre-processing
    - Data splitting
    - Model building strategies
    - Model performance and hyperparameter tuning
    - Results and final model selection (and web app presentation)
    - Discussion and conclusion
    - Your audience for this video is your *data science peers/technical audience* with the knowledge of data science methods and techniques.

For your technical report/notebook:

- Include a link to your team's GitHub repository.
- Include a clearly defined problem statement.
- Describe your analysis of the data.
- Include a detailed solution.
- You can include graphs and output tables, only if you use them in your discussion.
- This section is intended for a statistically literate/technical audience and must be written in a clear, organized fashion.
- You can organize the report into subsections such as:
    - Problem statement and justification for the proposed approach
    - EDA (graphical and non-graphical representations of relationships between the response variable and predictor variables)
    - Data wrangling and pre-processing (handling of missing values, outliers, correlated features, etc.)
    - Data splitting (training, validation, and test sets)
    - Model strategies (describing main research questions and appropriate analytics methods)
    - Validation and testing (model tuning and evaluation)
    - Results and final model selection (performance measures, etc.)
    - Discussion and conclusions (address the problem statement and suggestions that could go beyond the scope of the course)
    - Technical notebook code must be in Python

For your executive summary “Data Story” PowerPoint:

- Should be no longer than 6 slides.
- Introduce the problem statement.
- Discuss the solutions that were explored.
- Describe the conclusions of your data analysis to a non-technical audience.
- Should be intelligible to a person who does not know predictive modeling techniques.
    - Suppose you are submitting a report to a Director/VP/Executive-level audience who are not familiar with statistical terminology and predictive modeling methods. This can be seen as the executive summary/introduction of your report.
    - You will want to present your analysis as a “Data Story” instead of a technical report. You should articulate your final model results for everyone to understand with clear recommendations and actions.
    - Emphasis is on how you present your findings and recommendations vs. just the content of your slides.

Project Deliverables and Submission Format:

- Prepare a recorded video presentation of your project (slides are needed for the business audience) using a screencasting tool, such as Screencast-o-matic or Zoom to record your screen and provide a voice narration. Ensure that the sound quality of your video is good and each member presents an equal portion of the presentation. Export the video file to an **mp4 format**.
    - You may use any recording software you wish. You may want to utilize Screencast-o-matic which is integrated with Blackboard and linked below. You can access it using your

USDOne account login. View the Recording Video Presentation and Submission Guidelines for MS-ADS Students guide for additional recording instructions.

- Implement the project in Python using a Juypter notebook. Within your code, import all packages used for the project in the first cell, use code cells for code and comments, and use markdown cells for headings and descriptions. Generate **a PDF document** for submission with code, comments, and results within the notebook. **Attach this document as an Appendix to your technical report.**
    - Your technical report will consist of your Python notebook that contains detailed explanations of your work.
- Prepare and submit your Executive Summary and recommendation slides in PPT or PDF format.
- Submit the 15-20 minute presentation mp4 video file, technical report PDF document, and executive summary PowerPoint/PDF slides on the final team project submission page of Blackboard. You will use the naming convention Final Project Report-Team Number.pdf (e.g., Final Project Report-Team 1.pdf and Executive Summary Presentation-Team 1.pptx). Only one member of your team will need to submit these deliverables.
- You will submit the peer evaluation form individually.

To understand how your work will be assessed, view the assignment rubric .