ADS-506 Final Team Project Outline

Introduction:

As an integral part of this course, students are asked to find their own time series dataset, define a problem statement, and analyze and build a model with the dataset for the final project. You and your teammates will work closely together for the final project to find a dataset, clean and pre-process, and perform a time series forecast and/or analysis on the data using R or Python.

Deliverables:

There will be **two deliverables** for this project:

1. **Final Project Part 1:** A research paper in APA 7 format, geared towards a technical audience.
2. **Final Project Part 2:** A data story in the form of a PowerPoint Presentation including narration geared towards a non-technical audience.

It is required that you and your team use and practice GitHub as a code hosting platform for version control and collaboration during this project. It is also necessary to create and add a README.

Project Timeline:

- Module 2 (by the end of Week 2):
    - (by Day 2 of Week 2) The course instructor will group students into teams of two to three members. Blackboard, USD Email, or Slack can be used to communicate with prospective team members.
    - (by the end of Week 2) Each team should select and introduce a dataset (see potential dataset resources below) and the potential problem statement. The team representative will need to submit the "Final Team Project Start Form."
        - The problem statement should identify the gap between where the project or business currently is, and where they would like to be. This should include:
            - Description of the current state.
            - The business objective.
            - The success criteria.
            - The issue(s)
            - The threat(s) if applicable
- Module 3 (by the end of Week 3): Teams should:
    - Have a project introduction that includes a general overview of the project and project relevance.

    - Finalize their problem statement.
    - Complete a literature review pertaining to their project with a minimum of three relevant sources. Each literature review should include:
        - Background information that explains the working topic.
        - Overview of main points demonstrating gaps of knowledge, strengths, weaknesses, etc.
        - Key findings
- Module 4 (by the end of Week 4): Teams should:
    - Continue working on research papers.
    - Complete the performing of some basic exploratory data analysis to familiarize themselves with their dataset.
    - Complete the preprocessing (data cleaning, differencing, smoothing, and transforming if necessary) process that is unique to the need of their dataset.
    - Ensure that the tasks performed during EDA and pre-processing are written in the respective methodology paragraphs providing detailed information on the dataset characteristics and necessary pre-processing steps taken.
    - Begin writing their respective methodology paragraphs.
- Module 5 (by the end of Week 5): Teams should:
    - Build their time series models and defend their reasoning in their methodology section.
    - Conclude their methodology section with performance evaluation.
- Module 6 (by the end of Week 6): Teams should:
    - Provide a discussion within their papers.
    - Recommend steps to be taken that are perhaps beyond the scope of this class but may be practical or worth pursuing.
    - Provide a conclusion that summarizes the findings and any other notable details.
    - Finalize an APA reference page in addition to in-line citations.
    - Provide an appendix that contains code and information that would allow another data science student to repeat your analyses. Please do not paste raw code into the body of the paper.
    - Include your final project GitHub link in the report and the "Comment" session in the Blackboard.
    - Submit their written report (Final Team Project Part 1) in Word or PDF format.
- Module 7:
    - (End of Day 1): Instructor will send an announcement informing each team of the team whose project they're evaluating.
    - (End of Day 5) Each team will submit their completed Data Story Presentation (Final Team Project Part 2) to Blackboard and send it to their Peer Evaluators.

- (End of Day 7) Each team will submit their Final Team Project Peer Evaluation Forms. Only one representative from each team needs to submit this form to Blackboard.

*Beginning in Module 3, each team member will answer a short questionnaire on their team members to ensure that each team member is pulling their own weight throughout this project.

Possible Dataset Resources

- City of San Diego Datasets
- Federal Reserve Economic Data
- UCI Data Repository
- Google Dataset Search
- National Oceanic and Atmospheric Administration
- Healthdata.gov
- AWS Open Data
- Stock Market Data
- Economic Indicators
- World Bank DataBank
- USA.gov - Data and Statistics
- California Economic Indicators
- Bureau of Labor Statistics
- U.S. Census Bureau - Economic Indicators

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.

Part 1 Deliverable: Written Report

- Project introduction. This at a minimum should include:
  - General overview of your project
  - Project relevance
  - Motivation
- Creating your problem statement. The problem statement should identify the gap between where the project or business currently is, and where they would like to be. This should include:
  - Description of the current state.
  - The business objective.
  - The success criteria.
  - The issue(s)
  - The threat(s)
- Complete a literature review pertaining to their project. Projects should have a minimum of **three** relevant sources. Each literature review should include:

- Background information that explains the working topic.
- Overview of main points demonstrating gaps of knowledge, strengths, weaknesses, etc.
- Key findings

- All the steps should be explained, (including but not limited to):
  - Exploratory data analysis- any notable findings?
  - In data preparation and data cleaning steps:
    - What steps do you take to get it into a format amenable to analysis?
    - Are there missing variables or outliers?
    - Was there a need for smoothing, differencing, or transformation?
  - Which model was used and why?
  - Which evaluation metrics were used?
  - Describe the preliminary and interesting results for your main time series analysis.
- Discussions (what actions should be taken?),
  - Discussion of the results and what that means for the project.
  - Include possible extensions/next steps that are beyond the scope of the course but would be worth pursuing and/or practical to do.
- Conclusion
  - Conclusion of the entirety of the project including any notable findings.
- References and Appendices. Provide an APA reference page in addition to in-line citations. Give an appendix that contains code and information that would allow another data science student to repeat your analyses. Please do not paste raw code in the body of the paper.
- Prepare and submit your written report in Word or PDF format.

Final Team Project Part 1 Rubric

Criteria	**Achievement Level** **Meets or Exceeds Expectations**	**Achievement Level** **Approaches Expectations**	**Achievement Level** **Below Expectations**	**Achievement Level** **Inadequate Attempt**	**Achievement Level** **Non-Performance**
Background, Problem Definition, and Motivation 10%	The goal of the chosen problem is well-explained including a clear and succinct description of the background,	The goal of the chosen problem is set forth, there is a description of the background, and includes an adequate	There is mention of the chosen problem, a shallow description of the background, and motivation.	Does not include a clear description of the purpose of the problem and/or the issue and motivation.	Absence of background, problem definition, and project motivation.

	issue, and motivation. The objectives and success criteria for the problem are clearly explained including data goals, their success criteria, and a clear understanding of how it achieves the objectives.	description of the motivation. The objectives and success criteria for the problem are adequately explained including data goals and their success criteria.	Does not include a clear description of the objectives and success criteria for the problem and/or the data goals.	Does not include a clear description of the objectives and success criteria for the problem and/or the data goals.	
Data Exploration and Quality 15%	The EDA and Data Quality are very thorough and the process is thoroughly outlined. Data Quality portion provides a thorough explanation of the data profile. Including a clear and comprehensive data description for the relevant sources.	EDA and Data Quality portion provides a clear explanation of the data profile. Includes a clear and comprehensive data description for the relevant sources.	EDA and Data Quality portion provides a low level explanation of the data profile. Includes a data description that is lacking in detail of the relevant data sources.	Does not provide a clear explanation of the data profile. Does not include a clear data description.	EDA and Data quality section not present in paper.
Literature Review 15%	Utilizes three informative, relevant, and credible scholarly sources to support research findings. Demonstrates seamless integration into the body of the project.	Utilizes three informative, relevant, and credible scholarly sources to support research findings. Demonstrates seamless integration into the body of the project, however, one of the sources was of questionable	Utilizes two informative, relevant, and credible scholarly sources to support research findings. Demonstrates seamless integration into the body of the project.	Utilizes one informative, relevant, and credible scholarly source to support research findings. Demonstrates seamless integration into the body of the project.	Absence of literature review.

		relevance or quality.			
Modeling/ Model Evaluation 25%	Information on the data analysis process is clearly communicated. Chosen strategies listed clearly correspond with theoretical framework; excellent use of literature review/theoretical framework to help establish emerging themes. Model evaluation metrics are appropriate and well interpreted.	Information on the data analysis process is communicated. Chosen strategies listed correspond with theoretical framework; good use of literature review/ theoretical framework to help establish emerging themes. Model evaluation metrics are appropriate and well interpreted.	Information on the data analysis process is communicated but lacking in detail. Chosen strategies listed correspond with theoretical framework; use of literature review/ theoretical framework to help establish emerging themes. Model evaluation metrics are appropriate and interpreted.	Information on the data analysis process is not communicated well. Some of the chosen strategies listed do not correspond with the theoretical framework; some use literature reviews/theoretical framework to help establish emerging themes. and/or Model evaluation metrics are inappropriate or not correctly interpreted.	Modeling section is not present in paper.
Results/ Discussion / Recommended Next Steps 15%	Robust and comprehensive discussion and conclusion that discuss findings and clearly outline next steps.	Comprehensive discussion and conclusion that discuss findings and outline next steps.	Discussion and conclusion that discuss findings and outline next steps, but are lacking in minor details.	Discussion and conclusion that discuss findings and outline next steps, but are missing major details.	Discussion and conclusion section is not present.
GitHub Collaboration 5%	Evidence of project repository, a readme, teammate collaborations, and collaboration throughout the life cycle of the final project.	Evidence of project repository, teammate collaborations, and collaboration throughout the life cycle of the final project.	GitHub link was submitted but was missing collaborators or project repository.	Only a Github link was submitted.	No evidence of collaboration through GitHub.
Peer Evaluation Score 5%	Points ranging between: 100-90	Points ranging between: 89-80	Points ranging between: 79-70	Points ranging between: 69-60	Points ranging between: 59-0

Cohesiveness and Writing Mechanics 10%	Ties together information from all sources. Paper flows from one section to the next. Author's writing demonstrates an understanding of the relationship among material obtained from all sources. Demonstrates quality technical writing that is free of spelling and grammar errors. Sources appropriately cited in APA Style.	For the most part, ties together information from all sources. Paper flows with only some disjointedness. Author's writing demonstrates an understanding of the relationship across the material. Writing is clear and appropriate for a technical setting. Contains few spelling and/or grammar errors and/or needs some revision to meet expectations. Sources appropriately cited in APA Style but may have some formatting errors.	Does not tie together information. Paper does not flow and appears to be created from disparate issues. Writing contains several spelling and grammar errors and submission needs major revision to meet expectations. Sources not appropriately cited in APA Style.	Does not tie together information. Paper does not flow and appears to be created from disparate issues. Writing does not demonstrate understanding of the material. Writing contains several spelling and grammar errors and submission needs major revision to meet expectations. Sources not appropriately cited in APA Style.	Writing does not meet graduate level standards. and/or APA citations are not present or do not meet graduate level standards.

Part 2 Deliverable: Data Story PowerPoint Presentation with Narration

This assignment is due by **Day 5** of the learning week. When submitting your Data Story Presentation, you will also send it to the team your instructor has paired you with that will evaluate your project.

This week you will complete the last two components (Data Story (due on Day 5) and Peer Evaluation (Due on Day 7)) of your Final Team Project. Based on Part 1 of the Final Team Project, you will create a Data Story PowerPoint presentation that you will share to a business audience.

1. Prepare a recorded video presentation of your project (slides are needed for business audience) using a screencasting tool, such as Screencast-o-matic, PowerPoint, 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 within Blackboard and linked below. You can access it using your USDOne account login information. View the Recording Video Presentation and Submission Guidelines for MS-ADS Students (document available in Blackboard) guide for additional recording instructions.
2. In your research paper, you extensively covered the technical aspects of your time series analysis project. In the Data Story portion of your final project, you are tasked with creating a data-centric story that a non-technical audience will be able to follow (think Executives and Stakeholders).
3. Put your data and findings at the center of your story as you build your narrative off of how you tackled your business problem/question/task. Do not use any data science-specific jargon when narrating your data story and be sure to explain your story in plain English so your findings are accessible to anyone without a data science background. You are free to choose the style of PowerPoint presentation that you think best suits your project. Please keep in mind as a general rule of thumb that the pre-loaded generic PowerPoint templates are not recommended for professional presentations or for this project. Your presentation slides should be neat, well thought out, and visually appealing. Please see the rubric for additional details pertaining to the grading.
    - Your data story at a minimum needs to include:
        - Title
        - Introduction to your Team
        - Your task/question/problem and background
        - A minimum of two effective visualizations
        - Discussion on findings geared towards a non-technical audience (Think plain language)
        - Solution / Call to action and Next Steps
        - References
4. There is no minimum or maximum requirement on the number of slides, but keep in mind that a general rule of thumb is to spend about one minute per slide. All group members must participate in the creation and narration of this presentation.
    - This presentation should be at least 8 minutes in length and a maximum of 10 minutes. It is important to create clean and intentional visuals. Effective data stories pay close attention to details such as focusing on a compelling story, and carefully choosing colors and imagery.
5. Your team will submit the Final Team Project Peer Evaluation of another team's Data Story Presentation in Assignment 7.1 below.

Final Team Project Part 2 Rubric

Criteria	Achievement Level Meets or Exceeds Expectations	Achievement Level Approaches Expectations	Achievement Level Below Expectations	Achievement Level Inadequate Attempt	Achievement Level Non-Performance
Presentation contains the necessary slides and are formatted in a clean, aesthetically pleasing, and professional manner. 25%	Presentation slides contain the necessary components; the presentation deck is clean, well designed, and professional.	Presentation slides contain the necessary components; the presentation deck is clean, and professional.	Presentation is missing a component and/or the presentation deck is lacking in visual and professional appeal.	Presentation is missing two or more components and/or there is no apparent effort in creating a professional and aesthetically appealing deck.	Presentation slides are messy, not professional, and/or lacking in effort.
Presentation contains a compelling data story and fluidly describes the analysis and results using non-industry specific language. (Language geared towards a non- technical audience.) 25%	Contains a **compelling** data story that is fully comprehensible to a non-technical audience.	Contains a data story that is fully comprehensible to a non-technical audience.	Data story lacks components, explanation of the problem and/or analysis are overly technical.	Lack of a cohesive data story, and/or explanation of the analysis is overly technical and/or lacking major components.	Presentation's data story does not meet graduate level standards.
Presentation contains a minimum of 2 data visualizations that are clean and easily discernible. 25%	Presentation contains a minimum of two, very easily interpretable visualizations, that are **well explained**, engaging and visually appealing.	Presentation contains a minimum of two, easily interpretable visualizations, that are engaging and visually appealing.	Presentation contains one easily interpretable visualization, that is engaging and visually appealing. Or, the presentation contains the minimum amount of visualizations,	Presentation contains one visualization. Or, the presentation contains the minimum amount of visualizations, but they are poorly constructed and not visually appealing.	Visualizations are not included in the presentation.

			but they are poorly constructed and not visually appealing.		
Presenter speaks in a clear and professional manner, is persuasive and influences the audience in the allotted time. 25%	Presenter's narration is engaging with concise and articulate points while translating technical concepts into terms their audience will understand. Uses allotted time wisely and is respectful of the audience. Each team member has a speaking part. Presentation is between eight to ten minutes in length.	Presenter's narration is concise and articulates points. Technical concepts are presented. Some explanations are over- or under-explained to adequately convey the message. Each team member has a speaking part. Presentation is between eight to ten minutes in length.	Presenter's narration is disjointed, disorganized and unprepared. Does not use time effectively and/or message not adequately conveyed. and/or Some team member(s) do not have a speaking part. Presentation is over or under 8-10 minutes in length.	Presenter's narration is disjointed, disorganized and unprepared. Does not use time effectively and/or message not adequately conveyed. and/or Some team member(s) do not have a speaking part. Presentation is over or under 8-10 minutes in length.	No recorded audio.