Final Team Project Part 1: Written Report

This assignment is due on Day 7 of the learning week. Your written report should contain each of the parts outlined below. Prepare and submit your written report in Word or PDF format.

Project Introduction

This at a minimum should include:

- General overview of your project
- Project relevance
- Motivation

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)

Literature Review

- 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

Explanation of Steps

- All the steps should be explained, (including but not limited to):
    - Exploratory data analysis- any notable findings?
    - 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

- 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

- 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.
- Include an appendix that contains code and information that would allow another data science student to repeat your analyses.
    - **Note:** Please do not paste raw code in the body of the paper.

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

Review: Final Team Project Timeline and Deliverables

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 in 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 evaluations.
- **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 are 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.

Potential 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.

To view the full Final Team Project Parts 1 and 2 prompts and rubrics, review the Final Team Project Outline .