Final Project - Step 3 (25 Points)

PSTAT100: Data Science Concepts and Analysis

Due Date

The deadline for this step is **June 9, 2025**.

Instructions

In this step, you will conduct a thorough data science analysis based on the foundational work you've completed in Step 2. This analysis should be comprehensive, addressing your research questions and testing your hypotheses with suitable models. Additionally, you will prepare a professional report to showcase your findings, interpretations, and conclusions.

1 Step 3: Final Project Report

1.1 Data Analysis

- Based on your research questions, dataset characteristics, and hypotheses, choose an appropriate analytical approach. This may involve statistical modeling, hypothesis testing, or in-depth exploratory data analysis using visualizations.
- Clearly justify your chosen method(s) and explain how they address your research questions. Where applicable, discuss any assumptions made and their implications.

Possible Approaches Include:

Regression Analysis: To explore relationships between continuous response and explanatory variables.

Classification: For predicting categorical outcomes (e.g., logistic regression, decision trees).

Clustering: To identify natural groupings within the data.

Time Series Analysis: For analyzing data with temporal structure.

Hypothesis Testing: To statistically test assumptions about your data (e.g., t-tests, chi-square tests, ANOVA).

Exploratory Data Visualization: For uncovering patterns, trends, or anomalies through detailed graphical analysis.

1.2 Evaluation and Interpretation

- Use appropriate evaluation techniques based on your selected approach:
- For models, apply relevant performance metrics (e.g., RMSE, accuracy) and include diagnostics to assess validity and assumptions.
- For hypothesis testing, clearly state your null and alternative hypotheses, report p-values or confidence intervals, and interpret the results.
- For visualizations, provide thoughtful interpretation and explain how the visuals support your research findings.

1.3 Interpretation of Results:

Clearly interpret the results of your models, drawing connections back to your hypotheses and research questions. Discuss any significant patterns, anomalies, or insights found in your analysis.

1.4 Visualization and Communication of Findings:

Use clear and informative visualizations to support your analysis. Annotate visualizations and discuss their significance. Ensure all visualizations have captions, labels, and are well-documented.

2 Final Report

Your report should be a structured, professional document (in .pdf format) that clearly presents your analysis,findings, and conclusions.

- Limit the report to 10-12 pages, including graphics and tables. After completing your analysis, you will need to prepare a final report summarizing your work. This report should:

2.1 Required Sections:

- **Abstract**: Provide a brief summary of your analysis, key findings, and conclusions in a few sentences.
- **Introduction**: Provide an overview of your dataset, including its source, variables, and the key questions and hypothesis you are investigating.
- **Data Processing**: Describe the steps taken for data cleaning, handling missing values, and any necessary preprocessing. Mention any transformations or feature engineering applied to improve the model's performance.
- **Modeling Process**: Describe your approach to selecting and implementing models.
- **Results**: Present your key results, using visualizations and tables as needed.
- **Interpretation**: Discuss how your findings answer the research questions and address your hypotheses.
- **Visualization and Communication** Include well-labeled and informative visualizations, with captions explaining their relevance. Ensure clarity and professionalism in all graphical elements.
- **Conclusion and Recommendations** Summarize your key insights and provide actionable recommendations.Discuss any limitations of your analysis and suggest potential future directions.

Appendix (Optional)

Any additional figures, code snippets, or supplementary information.

2.2 Do:

- Provide clear and focused research questions and hypotheses that guide your analysis.
- Conduct thorough exploratory data analysis with detailed documentation of your findings.
- Use effective visualizations to enhance the understanding of your data.
- Present clear explanations of your models/ tests.
- Include relevant visualizations and interpretations of the regression results.
- Clearly document the steps you took in the regression analysis, including variable selection and transformations.
- Provide concise conclusions and actionable recommendations based on your regression model.
- Submit **ONLY** the .pdf file (10-12 pages) on Gradescope.

2.3 Do Not:

- Print any code.
- Print lists of data.
- Include irrelevant or redundant analyses.
- Ignore outliers or influential points in the dataset without explanation.
- Submit incomplete or poorly organized report.

3 Peer Review Process

- In addition to submitting the final report, you are required to evaluate your team members' performance and contributions using the **Final Project - Peer Review form on CANVAS**. This must be completed before the final project submission deadline.
- The peer review process is anonymous and should be conducted with honesty, responsibility, and fairness.

4 Deliverables and Submission Requirements

- Submit your report in PDF version by the due date (**June 9, 2025**).
- Ensure that your report is well-organized, with all sections clearly labeled and visualizations included.
- **Completed Peer Review** for each team member (submitted individually).

5 Additional Notes

- The final report is your opportunity to communicate your analysis and findings in detail. Make sure it reflects your understanding of the dataset, your regression techniques, and your diagnostic checks.