Data Science Capstone Topic Approval Form

Student Name: Shanikwa Haynes

Student ID: Click here to enter text.

Capstone Project Name: Predictive Analysis of State Correctional Spending: A Multi-Variable Regression Study of Recidivism, Reentry Investment, and Operational Cost Factors

Project Topic: Statistical Analysis of Correctional Spending Patterns and Recidivism Outcomes Across U.S. States

☒ **This project does not involve human subjects research and is exempt from WGU IRB review.**

Research Question: *Summarize one question or decision you will address by collecting…* What factors significantly predict total correctional spending across U.S. states, and how do recidivism rates, reentry investments, and operational cost components contribute to overall correctional expenditures?

> **Hypothesis**: This study will test the hypothesis that spending patterns in specific correctional categories (reentry investment, general operations, medical costs, and educational programs) significantly predict total state correctional spending, with reentry investments showing an inverse relationship to recidivism rates.
> **Null Hypothesis**- There is no significant relationship between recidivism rates, reentry investment spending, operational costs (general operations, medical, educational), and total correctional spending across U.S. states ($\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$).
> **Alternate Hypothesis**- At least one of the predictor variables (recidivism rates, reentry investment spending, general operations, medical costs, or educational program spending) is a significant predictor of total correctional spending across U.S. states (at least one $\beta \neq 0$).

Context: *Explain why the situation or question would benefit from a data analysis in fewer than 500 words.* State correctional systems face mounting pressure to reduce recidivism rates while managing increasing inmate populations and limited budget resources. With annual correctional spending exceeding $80 billion nationally, understanding the relationship between spending patterns and outcomes is critical for policy development and resource optimization. This analysis addresses a significant gap in correctional policy research by providing quantitative evidence of how different spending categories relate to both total expenditures and recidivism outcomes. The research will benefit state correctional administrators, policymakers, and budget analysts by identifying which spending strategies provide the best return on investment for reducing recidivism and improving public safety. Current policy debates about prison reform and rehabilitation funding lack comprehensive statistical analysis of spending effectiveness across multiple states. This study will provide evidence-based recommendations for optimal resource allocation in correctional systems, potentially influencing policy decisions that affect thousands of inmates and millions of taxpayer dollars annually.

Data: *Identify data you will need to collect that is relevant to the situation or question.* The analysis requires comprehensive correctional spending data and recidivism statistics from all 50 U.S. states, including: recidivism rates (percentage of released inmates returning to incarceration within three years), reentry investment spending (state funding for reentry programs and services), general operations costs (basic facility operations and staffing), medical costs (healthcare expenditures for inmates), educational program spending (vocational and academic training investments), and total correctional spending (combined expenditures across all categories). Data represents fiscal year spending amounts in millions of dollars and standardized recidivism measurement criteria.

> *If an existing data set will be used, describe the data set.* The dataset contains correctional spending and recidivism data for all 50 U.S. states with seven variables: State (categorical identifier), Recidivism_Rate (percentage), Reentry_Investment (millions USD), General_Operations (millions USD), Medical_Costs (millions USD), Educational_Programs (millions USD), and Total_Spending (millions USD). The data represents fiscal year expenditures and three-year recidivism rates, providing a comprehensive view of state correctional spending patterns and outcomes.

> *Explain who owns the data and why you are allowed to use this data for your capstone project.* This data is compiled from publicly available state correctional department reports and federal Bureau of Justice Statistics publications. The data consists of aggregate, non-personal information that is routinely published

by state governments for public transparency and accountability purposes. As publicly available government data, it can be freely used for academic research without restrictions. No proprietary or confidential information is involved, and all data sources are cited appropriately in academic publications.

Note: If you are using restricted information, please have the Third-Party Authorization Form signed by an authorized agent on behalf of the data owner. The data owner's legal name is required on the form.

Data Gathering: *The data-gathering methodology employs secondary data analysis of officially reported state correctional expenditures and recidivism statistics. Data is compiled from state departments of corrections annual reports, federal Bureau of Justice Statistics publications, and standardized state financial reporting systems. This methodology ensures data reliability through official government verification while enabling comprehensive cross-state comparisons using consistent measurement criteria and reporting periods.*

Data Analytics Tools and Techniques: *he analysis will employ multiple linear regression as the primary statistical technique, supplemented by descriptive statistics, correlation analysis, and comprehensive diagnostic testing. Additional techniques include exploratory data analysis with visualization (histograms, scatterplots, correlation matrices), assumption testing for linear regression (linearity, normality, homoscedasticity, independence), residual analysis for model validation, and effect size calculations for practical significance assessment. All analysis will be conducted using Python/R statistical software with supporting packages for data manipulation, visualization, and statistical modeling.*

Justification of Tools/Techniques: *Multiple linear regression is the most appropriate technique for this analysis because it directly addresses the research question about factors predicting total spending while allowing simultaneous examination of multiple predictors. The technique provides quantifiable relationships between predictors and outcomes, enables statistical inference through hypothesis testing, and offers interpretable coefficients for policy recommendations. Python/R is selected for its comprehensive statistical capabilities, extensive regression diagnostic tools, and ability to produce publication-quality visualizations. This combination provides the statistical rigor required for graduate-level research while ensuring reproducible, evidence-based results suitable for policy applications.*

Project Outcomes: *The key anticipated outcomes include: (1) A comprehensive statistical model explaining variance in state correctional spending with quantified relationships between spending categories and total expenditures, (2) Evidence-based recommendations for optimal correctional resource allocation based on statistical analysis of spending effectiveness, (3) Professional research report suitable for academic publication and policy application, (4) Complete Python/R analysis framework that can be replicated for ongoing correctional spending analysis, (5) Identification of spending patterns associated with lower recidivism rates, and (6) Statistical evidence to inform state budget allocation decisions and correctional policy development. Deliverables will include a formal research report, complete statistical analysis documentation, professional visualizations, and reproducible Python/R code for future analysis.*

Projected Project End Date: 08/17/2025

Sources: The only sources used were the official course materials from WGU. No outside sources were used.

Instructor Signature/Date:

☒ The research is exempt from an IRB Review.

☐ An IRB approval is in place (provide proof in appendix B).

Instructor's Approval Status: Approved

Date: 7/21/2025

Reviewed by:

Comments: Click here to enter text.