I. PROJECT OVERVIEW

Key user attributes: Gender, Race/Ethnicity, Parental Level of Education, Lunch, Test Preparation Course, Math Score, Reading Score, Writing Score

II. LIBRARIES AND DATA HANDLING

Libraries Used: Pandas, Seaborn, Matplotlib, SciPy,Statsmodels,Scikit-Learn

Data Loading and Preprocessing: Loading CSV file into a Pandas dataframe

III. DATA ANALYSIS TECHNIQUES

Descriptive Statistics: Descriptive Methods (e.g., df.head(), df.tail())

Inferential Statistics: T-test

Predictive Modeling: Linear Regression, Ridge,

Lasso, Random Forest, AdaBoost Regressor

Visualization Methods: Histograms, Bar Charts, Pie

Charts, Violin Plots

IV. KEY FINDINGS

Model Evaluation: Mean Squared Error (MSE) and R-squared (R²)

V. ADVANCED ANALYSIS

Feature Importance Analysis: Highlights most influential factors affecting student performance in math, reading,, and writing exam.

VI. MACHINE LEARNING IMPLEMENTATION

Linear Models: Linear Regression, Ridge Regression, Lasso Regression

Non-Linear Models: Random Forest Regressor, AdaBoost Regressor

VII. VISUAL INSIGHTS

Distribution of user attributes: Counts distribution of gender, race/ethnicity, parental level of education, lunch, and test preparation courses of each student. **Distribution of exam scores across attributes:** Correlation of math, reading and writing exam scores across individual user attributes

VIII. CONCLUSION

Summary of insights derived from the analysis, implications into related organizations.

APPENDIX

Datasets: Sample dataset of Student Performance in Exam used for data analysis

Code Snippets: Provided Python code used for loading, pre-processing, data analysis, and generating visualizations.

Google Colab Link:

https://colab.research.google.com/drive/15ZzTj8lywq 6qCMzLZO8XN6CWDuEXBzN1?usp=sharing

Github Website Link:

https://renzdxtr.github.io/CSST104-Final-Project/