Regression Analysis on Student Grades:

We are an educational consultancy company interested in understanding factors that influence student performance. We are looking for a regression model to predict the final grades of students based on a vast array of personal and socio-economic factors. Accurate predictions can help in identifying students who might need additional support. We have the data in a CSV format exported from Excel, but we would like this transferred to a local Postgres database. Additionally, we require a python function that can add a new student to the database.

Requests:

- Perform EDA with visualizations to assist in feature selection and engineering
- Convert the CSV to a Postgres database and create a python function to easily add new students to the Postgres database. Ensure the database schema has appropriate datatypes.
- Selection and/or engineer the feature set you will use for your model
- Select and train a model to predict the final grade of a student.
- Tune the model to get the best results possible
- Generate valid metrics to evaluate your model

Additional Requests (Bonus):

- Create another model that predicts final grades **without** using any of the previous period grades. This would be immensely helpful to the company to aid in predicting student performance before they fall behind
- Write a full Data Analysis report on the statistics gleaned from the dataset
- Include documentation used to plan out this project and its timeline.

Requirements:

- All code shall be written in python or SQL
- All code shall be managed via git
- The repository shall be named '<first_name>-<last_name>-aml-student-regression' (all lowercase, i.e. Tom Cruise's repository name is 'tom-cruise-aml-capstone')
- Only libraries inherent to python or listed below can be used
- The Postgres database should be managed via docker
- Only data from the given dataset will be used

Allowed Dependencies

- scikit-learn
- numpy
- scipy

- pandas
- matplotlib
- seaborn

This capstone is due **1/22/2024**. Ensure you use your time to optimize the performance of each step in creating your model. The dataset and dataset description will be provided in 'assessment_1.zip'