# Machine Learning Assignment Summary

This notebook presents a structured machine learning workflow applied to the StudentsPerformance.csv dataset. The project begins with importing essential libraries such as pandas, numpy, seaborn, matplotlib, and several modules from scikit-learn for data preprocessing, modeling, and evaluation.  
  
1. Data Loading and Preprocessing  
The dataset is loaded and inspected for missing values, which are found to be absent. Categorical variables including 'gender', 'race/ethnicity', 'parental level of education', 'lunch', and 'test preparation course' are encoded using LabelEncoder for simplicity. The features are then separated from the target variable, which is the 'math score'. All features are standardized using StandardScaler.  
  
2. Model Building: Linear and Polynomial Regression  
Two regression models are developed to predict students' math scores:  
- Linear Regression: A baseline model is trained on the original standardized features.  
- Polynomial Regression: A polynomial transformation (degree 2) of the features is applied using PolynomialFeatures, followed by fitting a linear model.  
  
3. Model Evaluation  
The models are evaluated on test data using standard metrics:  
- Mean Absolute Error (MAE)  
- Mean Squared Error (MSE)  
- Root Mean Squared Error (RMSE)  
- R² Score  
  
Results show that the polynomial regression model outperforms the linear model, as indicated by lower error values and a higher R² score.  
  
4. Visualizations  
The notebook includes visualizations for:  
- Feature distributions  
- Correlation heatmaps  
- Predicted vs actual math scores (for both models)  
  
These plots help understand the data structure and the effectiveness of the models.