**Title: Predicting Student Performance using Linear Regression**

**Objective:**

The objective of this project is to analyse the relationship between various factors and student academic performance, and to build a predictive model using linear regression. We will utilize data science techniques to explore, preprocess, visualize, and model data, focusing on predicting student grades based on features such as study time, previous exam scores, and personal characteristics.

**Dataset:**

We will use a dataset containing information about students, including attributes such as study time, previous exam scores, family background, and final grades. The dataset can be found at **UCI Machine Learning Repository**.

**Steps:**

Data Collection: Download the dataset from the provided link or any other reliable source.

**Data Preprocessing:**

Load the dataset into a DataFrame (using pandas).

Handle missing values and perform any necessary data cleaning.

Encode categorical variables if needed.

Split the data into training and testing sets.

**Exploratory Data Analysis (EDA):**

Perform descriptive statistics to understand the distribution of data.

Visualize relationships between features and the target variable (grades) using matplotlib. Examples include scatter plots, histograms, and box plots.

**Feature Selection:**

Identify relevant features that may influence student performance. This can be done through correlation analysis or domain knowledge.

Select features that have the most significant impact on grades.

**Model Building:**

Implement linear regression using libraries like scikit-learn.

Train the model on the training dataset.

**Model Evaluation:**

Evaluate the performance of the model using metrics such as Mean Squared Error (MSE), R-squared, and visualizations (e.g., residual plots).

Test the model on the testing dataset and assess its generalization performance.

**Results Visualization:**

Visualize the predicted grades versus the actual grades using scatter plots.

Plot the regression line on the scatter plot to show the model's fit.

**Conclusion:**

Summarize findings and insights from the analysis.

Discuss the model's strengths, limitations, and potential areas for improvement.

**Tools and Libraries:**

Python programming language

pandas for data manipulation

matplotlib for data visualization

scikit-learn for machine learning

**Deliverables:**

Jupyter Notebook documenting the entire project workflow, including code and explanations.

Presentation slides summarizing key findings, insights, and the predictive model.

Report detailing the methodology, results, and conclusions.