**Assignment**

The purpose of this project is to delve into student data to uncover insights that could help understand and predict student success across different academic periods. As a contributor, you will explore the relationship between student grades and a myriad of demographic, social, and school-related factors. Your analysis will directly impact our ability to identify and address key influences on student performance.

**Your Tasks:**

1. **Create Data Visualizations:**
   * Generate histograms to observe the distribution of grades (G1, G2, G3) and other numerical factors like age and study time.
   * Construct box plots to spot outliers and understand the spread of the data.
   * Use bar charts to compare the average grades across different categories such as gender, parental education level, and internet access.
   * Your visualizations should help highlight trends and patterns that may influence student performance.
2. **Perform Statistical Tests:**
   * Conduct Chi-Square Tests to investigate the association between categorical variables (e.g., gender, internet access) and student grades. For example, is there a statistical difference in grades between students with different levels of parental education?
   * Apply t-tests to compare the mean grades between two different groups, such as students from urban versus rural areas.
   * Use regression analysis to predict final grades based on various factors like study time and past failures.
   * Document your findings, interpret the p-values, and discuss the statistical significance of your results.

**Objectives:**

* **Correlation Analysis:** Determine which factors are most strongly correlated with student grades.
* **Predictive Modeling:** Build a model that can predict a student's final grade based on their background and school-related activities.
* **Insight Generation:** Provide actionable insights for schools to help improve student outcomes based on your findings.

**Tools You Might Need:**

* **For Visualization:** Matplotlib, Seaborn, or any other Python library that you are comfortable with.
* **For Statistical Analysis:** SciPy for conducting statistical tests, and statsmodels or Scikit-learn for any predictive modeling.

**Data Description**

The dataset consists of two files, **student-mat.csv** and **student-por.csv**, which include various features that describe student demographics, family background, school-related factors, and academic performance. Below is a detailed breakdown of each variable:

**Student Information**

* **school** (*Categorical*): The school attended by the student ('GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira).
* **sex** (*Binary*): Student's gender ('F' - Female, 'M' - Male).
* **age** (*Integer*): Student's age (ranging from 15 to 22 years old).
* **address** (*Categorical*): Student's home address type ('U' - Urban, 'R' - Rural).

**Family Background**

* **famsize** (*Categorical*): Family size ('LE3' - Less or equal to 3, 'GT3' - Greater than 3).
* **Pstatus** (*Categorical*): Parent's cohabitation status ('T' - Living together, 'A' - Apart).
* **Medu** (*Integer*): Mother's education level (0 - None, 1 - Primary, 2 - 5th to 9th grade, 3 - Secondary, 4 - Higher education).
* **Fedu** (*Integer*): Father's education level (0 - None, 1 - Primary, 2 - 5th to 9th grade, 3 - Secondary, 4 - Higher education).
* **Mjob** (*Categorical*): Mother's job ('teacher', 'health', 'services', 'at\_home', 'other').
* **Fjob** (*Categorical*): Father's job ('teacher', 'health', 'services', 'at\_home', 'other').
* **guardian** (*Categorical*): Student’s guardian ('mother', 'father', 'other').

**Academic and School-Related Factors**

* **reason** (*Categorical*): Reason for choosing the school ('home', 'reputation', 'course', 'other').
* **traveltime** (*Integer*): Home to school travel time (1 - <15 min, 2 - 15-30 min, 3 - 30 min-1 hour, 4 - >1 hour).
* **studytime** (*Integer*): Weekly study time (1 - <2 hours, 2 - 2-5 hours, 3 - 5-10 hours, 4 - >10 hours).
* **failures** (*Integer*): Number of past class failures (numeric: 1 if 1<=n<3, else 4).

**Educational Support & Extracurricular Activities**

* **schoolsup** (*Binary*): Extra educational support ('yes' or 'no').
* **famsup** (*Binary*): Family educational support ('yes' or 'no').
* **paid** (*Binary*): Extra paid classes for Math/Portuguese ('yes' or 'no').
* **activities** (*Binary*): Participation in extracurricular activities ('yes' or 'no').
* **nursery** (*Binary*): Attended nursery school ('yes' or 'no').
* **higher** (*Binary*): Aspiration to take higher education ('yes' or 'no').
* **internet** (*Binary*): Internet access at home ('yes' or 'no').
* **romantic** (*Binary*): In a romantic relationship ('yes' or 'no').

**Social and Lifestyle Factors**

* **famrel** (*Integer*): Quality of family relationships (1 - Very bad to 5 - Excellent).
* **freetime** (*Integer*): Free time after school (1 - Very low to 5 - Very high).
* **goout** (*Integer*): Going out with friends (1 - Very low to 5 - Very high).
* **Dalc** (*Integer*): Workday alcohol consumption (1 - Very low to 5 - Very high).
* **Walc** (*Integer*): Weekend alcohol consumption (1 - Very low to 5 - Very high).
* **health** (*Integer*): Current health status (1 - Very bad to 5 - Very good).
* **absences** (*Integer*): Number of school absences (0 to 93).

**Academic Performance (Target Variables)**

* **G1** (*Integer*): First period grade (0 to 20).
* **G2** (*Integer*): Second period grade (0 to 20).
* **G3** (*Integer*): Final grade (0 to 20, target variable).