**PROJECT PROPOSAL**

**Student Habits and Academic Performance**

**Project Title:**

**The relationship between Student Habits and Academic Performance**

**Purpose and Outcome:**

* **Purpose:**

This project aims to:

* Explore how student habits affect exam scores.
* Identify key influencing factors
* Build a prediction model for exam scores.
* Provide actionable recommendations to help students improve their results.
* **Outcome**
* Understanding the relationships between student habits and exam score.
* Identify the factors that have the greatest impact on the exam score.
* Prediction model for exam score based on student habits.
* Actionable recommendations for improving exam score.

**Dataset:**

* **Source:** Public dataset from Kaggle (Student Habits vs Academic Performance): https://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance/data.
* **Description:** This dataset explores the relationship between student habits and exam score. It contains multiple variables related to study habits, health habits, and enviromental factors.
* **Structure:**
  + exam\_score: The final exam score of each student.
  + student\_id: A unique identifier assigned to each student.
  + age: The age of the student.
  + gender: Indicates the student’s gender.
  + study\_hours\_per\_day: the number of hours spent studying per day.
  + social\_media\_hours: the number of hours spent on social media daily.
  + netflix\_hours: the number of hours spent watching Netflix per day.
  + part\_time\_job: Indicates whether the student has a part-time job.
  + attendance\_percentage: Percentage of classes attended by the student.
  + sleep\_hours: the number of hours of sleep per night.
  + diet\_quality: Quality of the student's diet.
  + exercise\_frequency: Number of exercise sessions per week.
  + parental\_education\_level: Highest education level attained by parents.
  + internet\_quality: Describes the quality of the student's internet connection.
  + mental\_health\_rating: A self-reported mental health score.
  + extracurricular\_participation: Indicates if the student is involved in extracurricular activities.

**Initial Analysis Plan:**

* **Data Cleaning:** Data profiling, removing duplicates, handling missing values, ensuring consistency, fixing incorrect data types, identifying and handling outliers.
* **EDA:** Generate summary statistics, visualize distributions and explore relationships between variables.
* **Analysis:** Perform correlation analysis, build and evaluate a multiple linear regression model, and identify key factors contributing to exam score.
* **Visualization:** Create visualizations such as histograms, heatmaps, … to support the analysis.
* **Data Storytelling:** Present the findings, actionable recommendations for improving exam score.