**Problem statement**

**The primary goal of gathering data on factors affecting student performance is to discover the important aspects that influence academic outcomes. Understanding these elements allows schools and universities to devise focused measures to increase student achievement. This data tries to demonstrate trends and relationships between many aspects, such as demographics, study habits, and socioeconomic position, and how these affects academic success. By examining these indicators, we can address issues such as high dropout rates and poor grades and identify where resources should be directed.**

***Describing the Dataset and the Data Types that was used***

**Hours\_Studied (Integer): It suggests the number of hours a student devotes towards his books. It is an integer value that proves useful in a study conducted to determine the correlation between the amount of study time and the results achieved.**

**Attendance (Integer): Attends to the student’s attendance level in percentage form. This integer value is important for seeing any relationship that is present between good class attendance and better exam results.**

**Parental\_Involvement (String/Categorical): Explains how much the parent is involved in the student’s academic life, which could be low, medium or high. Known as a categorical variable, it assists in evaluating the effects of a particular factor namely, parental support on the performance of a particular student.**

**Access\_to\_Resources (String/Categorical): Refers to the availability of learning resources students and any other requirements in academics like library, books, computers etc., in the form of Low, Medium or High. This can be used in order to assess whether or not the availability of resources affects the learning results.**

**Extracurricular\_Activities (Boolean/String): Indicates if the student is involved in extra co-curricular activities, coded as ‘1’ for Yes, and ‘0’ for No. This binary variable assists in determining whether or not engaging in other activities apart from academics have an impact on performance.**

**Sleep\_Hours (Integer): Hours of sleep which the student spends per day. It helps to reduce the bias while analysing the relationship between sleep patterns and concentration as well as the academic performance.**

**Previous\_Scores (Integer): This brings out the student’s scores from previous assessments. The aforesaid historical data point can also be used in modelling current performance trends and noting areas of concern.**

**Motivation\_Level (String/Categorical): Describes the student’s motivation status; it could be low, moderate or high. Such qualitative data used in the research aids in explaining the effect of motivation on learning and examination performance.**

**Internet\_Access (Boolean/String): Indicates whether the student has Internet connection or not, by a ‘Yes’ or a ‘No’. This factor is particularly vital today that education has turned to rely on internet especially in accessing learning materials.**

**Tutoring\_Sessions (Integer): A variable illustrating the total amount of tutoring sessions the student participated in. This integer value assists in analysing whether extra academic aid has a positive impact and consequently raise the performance.**

**Family\_Income (String/Categorical): Defines the type of income of student’s family as low, medium or high-income group. It assists in finding out whether or not economic status has a role to play in academic achievement.**

**Teacher\_Quality (String/Categorical): Explains the student’s perceived quality of the teachers he or she gets, which could take the value of ‘Low’, ‘Medium’, or ‘High’. This factor assists in checking whether there is a connection between the quality of teachers and achievement of students.**

**School\_Type (String/Categorical): Tells the type of school the student goes to, ‘public’ or ‘Private’. Analysing this variable, we can get some insights about how students perform differently in various educational contexts.**

**Peer\_Influence (String/Categorical): Explains the extent, nature and intensity of the impact of peers on the student as ‘Positive’, ‘Negative’, or ‘Neutral’. It assists to evaluate the effects of social factors towards the academic conduct.**

**Physical\_Activity (Integer): Assesses the extent to which the student involves himself/ herself in physical exercise usually on a rating scale (such as 1-5). Lack of personal care results in damages in physical health that affect concentration and overall well-being.**

**Learning\_Disabilities (Boolean/String): The variable identifies whether the student has any diagnosed learning disabilities, and it has ‘Yes’ and ‘No’ as possible values. This information can be used to design programs to support the students as well as make provisions for barrier-free facilities.**

**Parental\_Education\_Level (String/Categorical): Shows the degrees of the student’s parents that education has offered them (for example ‘High School’, ‘College’, ‘Postgraduate’). This is useful especially when you want to analyse the background socio-economic environment impact on students’ performance.**

**Distance\_from\_Home (String/Categorical): Holds the values indicating the distance from the student home to school for example, ‘Near’, ‘Moderate’, ‘Far’. This may affect the timing of the day-to-day mobility and the energy that is used up in midst of the day.**

**Gender (String/Categorical): Enables the determination of sex of the student whether Male or female. This categorical variable enables one to ascertain performance trends from the gender perspective.**

**Exam\_Score (Integer): In this case, it depicts the student’s performance in the final exam. This is a way of measuring academic performance as a primitive indicator and it is an integer so that scoring can be easily done.**