Capstone III Project Proposal

1. *What dataset or datasets do you plan to use? What are the features, rows, and data types of each?*

The dataset I’m planning to use is a CSV file from [Kaggle](https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression). It contains data pertaining to student performance in accordance with various factors The columns and variable types are as follows:

|  |  |  |
| --- | --- | --- |
| Variable | Quantitative or Qualitative? | Variable Type |
| Hours Studied | Quantitative | Continuous |
| Previous Scores | Quantitative | Discrete |
| Extracurricular Activities | Qualitative | Binary |
| Sleep Hours | Quantitative | Continuous |
| Sample Question Papers Practiced | Quantitative | Continuous |
| Performance Index | Quantitative | Discrete |

1. *What research or business questions do you want to answer?*

There are two questions I intend to answer based on the data. The first is: Is there a significant difference in the performance of students based on their involvement in extracurricular activities? The second is: What is the relationship between sleep hours and student performance?

1. *What are your hypotheses going in?*

My first hypothesis is that there will be a significant difference of performance between students who are involved with extracurricular activities and those who are not.

My second hypothesis is that there will be a positive correlation between sleep hours and student performance.

1. *How will you use your data to test your hypotheses?*

I will be performing an independent samples t-test to conclude if there is a significant difference between the performance of students who are involved with extracurricular activities and those who are not. I will be using the Pearson correlation coefficient to determine the relationship between sleep hours and performance

1. *Who will find your findings valuable and how will they use them?*

I imagine students, parents and related professionals would find my findings to be valuable, and they could be used to determine both how much sleep could be impacting their performance / whether or not to participate in extracurricular activities.