Capstone III Project Proposal – Exemplar

*Students: This is an example of a project proposal for Capstone III. It is intended for guidance purposes only. You’ll need to come up with your own proposal, so be as creative as you want!*

Following are the questions you’ve been asked to address:

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

The dataset that I plan to use is a CSV file that [comes from Kaggle.](https://www.kaggle.com/spscientist/students-performance-in-exams) It contains test scores from high school students in the U.S. Below is each column along with its variable type.

|  |  |  |
| --- | --- | --- |
| Variable | Quantitative or qualitative? | Variable type |
| Gender | Qualitative | Binary |
| race/ethnicity | Qualitative | Nominal |
| parental level of education | Qualitative | Ordinal |
| lunch (type) | Qualitative | Binary |
| test preparation score | Qualitative | Binary |
| math score | Quantitative | Continuous |
| reading score | Quantitative | Continuous |
| writing score | Quantitative | Continuous |

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

I would like to answer two questions using the data. The first is to see if there is any correlation between math, reading, and writing scores. I will check for each of these pairwise relationships (math and reading, reading and writing, math and writing).

The second is to see if there’s a significant difference between scores for those students who had a standard lunch versus a free/reduced lunch.

1. *What are your hypotheses going in?*

My first hypothesis is that the student scores are correlated. This means that students who score well in one subject are likely to score well in another subject. I would expect correlation coefficients of at least 0.5 to confirm this.

My second hypothesis is that there is no difference between scores depending on the type of lunch provided. I will test this at the 95% significance level.

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

I will use scatterplots and the Pearson correlation coefficient to check correlation. I will use the independent samples t-test to check for average score differences.

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

The local school district will be interested in knowing if students who score well do so across subjects, and they’ll also be interested in knowing if the type of lunch being served has an impact on performance. Knowing these things will be helpful in targeting resources toward improving performance for all students.