1. Dataset

The dataset contains Exam Scores and other information for students at a public school.

1.1 Sample Data

Column Label

Gender	Ethnicity	Age	Parental Level of Education	Test Preparation Course	Math Score	Science Score	English Score
male	Chinese	16	high school	none	47	77	79
male	Malay	13	high school	none	41	55	70
female	Chinese	14	diploma	none	69	50	64
female	Chinese	15	high school	completed	48	56	73
female	Chinese	13	undergraduate degree	none	43	46	61

1.2 Attributes for the dataset

1. **Gender**: Student's gender (binary: female or male)

2. **Age**: Student's age (numeric: from 13 to 17)

3. **Ethnicity**: Student's ethnicity (nominal: Malay, Chinese, Indian , Others)

3. **Parental Level of Education**: Highest education of student's parents (nominal: high school, diploma, undergraduate degree, postgraduate degree)

4. **Test Preparation Course**: (binary: none, completed)

5. **Math Score**: Student's math score (numeric: from 0 to 100)

6. **Science Score**: Student's science score (numeric: from 0 to 100)

7. **English Score**: Student's English score (numeric: from 0 to 100)

1.3 What we want to explore:

1. How effective is the test preparation course?
2. Which major factors contribute to test outcomes?
3. What would be the best way to improve student scores on each test?
4. What patterns and interactions in the data can you find?

1.4 Dataset to download:

	**Download a Small Sample** A .csv file with 10 results as a sample set (n = 10)
	**Download a Large Sample** A .csv file with 1000 results as a sample set (n = 1000)