**Assignment Problems: Day 4**

The following problems require access to the file named ‘Dataset\_Day4.csv’, provided with this assignment.

About the data,

|  |
| --- |
| Grades from a Chemical Engineering course at McMaster University. |
| The recorded values are the average of sub-components: e.g the Tutorial variable is the average of all tutorials, the Final exam variable is the average of all questions in the final, written exam.  The Prefix column is the year in which the student first enrolls at the university and is a crude approximation of the student's age (maturity).  This particular course permitted students to work in groups for assignments, tutorials and the take-home exam. The groups were self-selected, and varied during the semester.  Of interest is whether the assignments, tutorials, midterms or take-home exam are a good predictor of the student's performance in the final exam. Also, does the Prefix variable show any promise as a prediction variable? |

Problems to solve,

1. Find and treat all the missing values. Rows or columns with missing values must not be dropped.
2. Make an in-depth report on the variables using descriptive statistics and also portray the relationship between each variables using visualizations.