Problem Statement

Given to you is the information about two different variants of a product. ***"ProductA.csv"*** has information of ***type A variant*** and ***"ProductB.csv"*** is about **type B variant**. For each of the variant, there are values for several characteristics and the target is the grade of the product which is a score between 0 and 10. This problem has two parts

a. Predict the target for each of the product variant.
b. Given the characteristics, identify the variant of the product

1. Understand the data through visualizations
    a. You may use ggplots or tableau (which would be covered in the next classes) to visualize the data and provide any insights
    b. Submit your visualizations by 4th of May 2018
2. You may try different algorithms as and when you learn them in this module (CSE 7305c).
3. You must submit both R and Python codes of your solutions at the end of the module. The deadline for submissions will be mentioned later.
4. You may use both classification and regression algorithms depending on how you see the target. Understand well if there is a class imbalance when you go with classification approaches.
5. Which model gives you the best result? While testing this you may use cross validation approaches