In these first two parts, you will create a framework to scope out data science projects. This framework will provide you with a guide to develop a well-articulated problem statement and analysis plan that will be robust and reproducible. Before completing an analysis, it is critical to understand your data. You will need to identify all the biases and variables in your model in order to accurately assess the strengths and limitations of your analysis and predictions.

Dataset

We'll be using the same dataset as UCLA's Logistic Regression in R tutorial to explore logistic regression in Python, as explained in yhat's blog. This is an excellent resource for using logistic regression and summary statistics to explore a relevant dataset. Our goal will be to identify the various factors that may influence admission into graduate school. It contains four variables:

'admit' is a binary variable. It indicates whether or not a candidate was admitted (admit =1) or not (admit= 0)

'gre' is GRE score

'gpa' stands for Grade Point Average

'prestige' is the rank of an applicant's undergraduate alma mater, with 1 being the highest and 4 the lowest