The dataset contains 500 inputs, where every row is an Indian student that applied to a master program at the USA. Every row has nine parameters which are considered for the admissions comitee of the graduate schools. The parameters included are : 1. GRE Scores ( out of 340 ) 2. TOEFL Scores ( out of 120 ) 3. University Rating ( out of 5 ) 4. Statement of Purpose and Letter of Recommendation Strength ( out of 5 ) 5. Undergraduate GPA ( out of 10 ) 6. Research Experience (either 0 or 1 ) 7. Chance of Admit (ranging from 0 to 1 ).

This dataset has a calculated parameter -> `Chance of admit`. This parameter has been calculated by the owner of the dataset. Unfortunately, the information regarding if the applicant was admitted or not is not accessible. Also, it seems that this dataset only takes one application per subject, when it's a known fact that rarely one applicant applies to only one master program. Additional data from individuals could be useful, despite usually universities do not ask for age and gender information, this data could be obtained looking the amount of years passed from graduating and the name of the applicant. This information could be useful data to add in the future.

The boxplots let us visualize that numerical predictors have a normal distribution. Even the ranting predictors, despite of not having a strictly normal distribution they don’t present any obvious anomaly.

It can be observed that SOP and LOR tend to have a left asymmetry, in other words, is more common to have a good SOP and LOR than having a terrible one. This has sense given that students are not usually asking recommendation letters to advisors that don’t have a good concept of them. Also, students tend to effort to write adequate SOP.

We could observe that our predicted value ‘Chance of admit’ has strong correlation with several of ours predictors. This is a hint that these predictors could be adequate to predict admission to master program. The predictor most correlated with the outcome is the CGPA, with a positive correlation of $R^2=0.88$. It is necessary to appoint that CGPA also highly correlates wit GRE and TOEFL scores, this should be taken in count when we build a model given the possibility of collinearity between predictors.

CGPA and TOEFL looks that they could benefit from a quadratic transformation, and GRE looks something similar to a logarithmic transformation.