Project EDA

Anna Li, Jess Eng, Seth Billiau

Due: November 15, 2019

A Brief Description of the Dataset

Our dataset was exported from the Sports Reference College Basketball website. We extracted basic and advanced team statistics for the 330 NCAA Division 1 teams in the 2007-2008, 2009-2010 and 2017-2018 college basketball seasons. We will team statistics from the 2007-2008 (and possibly the 2008-2009) basketball seasons as our training set with the 2008-2009 W-L% as our response variable. This training set will be split into training and validation datasets to be used for model selection. The 2017-2018 dataset (and possibly 2019-2020) will be used as our test set. The accuracy of our model on this dataset will be unknown until the completion of the 2019-2020 season.