Project Proposal

A detailed description of the dataset addressing:  – Where is it from?  –

*This data set was compiled by the NCAA (the National Collegiate Athletic Association). It is an attempt to compare academic information about college student- athletes across both other schools and sports. The NCAA created the Academic Progress Rate to ensure sports teams were being held accountable for their academics in addition to sports. The Academic Progress Rate is a composite score based on student-athlete eligibility, retention and graduation. The higher the APR score indicates the more successful teams in the classroom. The NCAA uses the APR score to punish teams as well for teams that aren’t performing to acceptable academic standards.*

How many rows/columns are in this dataset?  –

* *There are 6521 rows and 57 columns. However, we plan to cross-reference this data with the college scorecard~~, as well as information about each team and college’s overall athletics’ performance[ie revenue, total number of championships, home game attendance~~].*
* What information does it contain?
* This contains:
* SCHOOL\_ID
* SCHOOL\_NAME
* SCHOOL\_TYPE
* ACADEMIC\_YEAR
* SPORT\_CODE
* SPORT\_NAME
* NCAA\_DIVISION
* NCAA\_SUBDIVISION
* NCAA\_CONFERENCE
* FOURYEAR\_ATHLETES
* FOURYEAR\_SCORE
* FOURYEAR\_ELIGIBILITY
* FOURYEAR\_RETENTION
* 20XX\_ATHLETES
* 20XX\_SCORE
* 20XX\_ELIGIBILITY
* 20XX\_RETENTION

A description of the question(s) you hope to answer with the dataset:

* + - * *Which sport or school has the highest academic score?*
      * *Which schools have improved or worsened the most?*
      * *Is there a difference between public and private institutions?*
      * *Smartest sport across schools?*
      * *Highest academic progress conference?*
      * *Do men’s or women’s teams perform better across the whole?*
      * *Does geographic location indicate* retention rates?
      * *How does the student body compare to the athletic student body?*
      * *Predict retention rates using multi linear regression (using stepwise regression to determine which variables are significant)*
      * *Does tuition predict retention?*