**Group :** College\_Score\_Card\_E139-Fall2015

Github: *Github:* [*https://github.com/aspdave/College\_Score\_Card\_E139-Fall2015*](https://github.com/aspdave/College_Score_Card_E139-Fall2015)

**Members:** Himanshu Dave [hdave@g.harvard.edu](mailto:hdave@g.harvard.edu) Phone: 732 835 8129

Michael [michael83@gmail.com](mailto:michael83@gmail.com)

Sway

Bin

**Project Outline** : Group is set out to establish on true cost of college education based on the College score card dataset ( <https://collegescorecard.ed.gov/data/>) . Project outcomes is create the Statistical model and provide toll to compare the colleges across based on predictor . Insights from this data analysis  can help prospective students to data based decision . Our investigation is limited by Federal dataset available online. Time series is from 1996 -2013 USA colleges.

**Accomplishment as of today** :

Project group is formed , kick off meeting is done. group ‘s private share folder is created , group blog is created to share ideas, GitHub code share is established. Initial college score card dataset is data set is downloaded, examine the dataset . Investigated data set for possible predictor variable and for response variable.

**Challenges ahead** : Defining the “Quality “ and “true” cost factor of college is real challenge . There is wide variety of variables to choose from. We do expect close to normal distribution due to large collection , there are good chances of outliers . we do need to check assumption and limitation solely based on data contain with dataset. Cohort definition as imperfect and varies by each metrics.

Model do not fit all of the population as students tends to have different Goals and individual expectation that varied widely.