**Capstone Project Ideas**

University student retention rates

Higher education is simultaneously expensive and critical to future economic success. One major impediment to future success is beginning an education, incurring significant debt, and then failing to complete a degree. I aim to examine retention rates across universities using the College Scorecard database (<https://collegescorecard.ed.gov/data/>). Moreover, I aim to compare like to like when examining schools. That is, I intend to group schools by SAT/ACT scores, first generation student percentages, family income values, and other factors that may impact retention rates. I hypothesize that factors such as factors such as the degree type offered, organization of the school (e.g., public or private), and any special mission of the school will have an impact on the rate of student retention.

Lending Club loan risk assessment

Access to credit is an important factor in a modern economy, as is accurately assessing the risk associated with borrowers. I aim to use the Lending Club loan data (https://www.lendingclub.com/info/download-data.action) to examine loan status (current, late, etc.) as a function of various attributes of the borrower. Specifically, I hypothesize that loan status will be heavily impacted by current length of employment and the frequency of loan inquiries.

Lending Club loan prediction

Beyond loan risk, the Lending Club data offers insight into the factors that may make a potential borrower too risky to lend to. These factors include the debt-to-income ratio, length of employment, loan purpose, and loan size. Additionally, there are geographic factors that may alter the calculus in riskiness. Here, I aim to predict whether or not a loan will be rejected based on the attributes above. Specifically, I hypothesize that geographic factors will play a substantial role in evaluating loan risk.